

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problems Mailbox.**

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
(c) 1993 - 2003 Compu

Search time 70 Seconds
(without 31 months)

612.953 Million cell updates/sec

.....LFEFRISRNGVVDPLTVLK 322

dues

imeters: 908470

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

/geneseqp-emb1/AA1980.DAT *	/geneseqp-emb1/AA1981.DAT *	/geneseqp-emb1/AA1982.DAT *	/geneseqp-emb1/AA1983.DAT *	/geneseqp-emb1/AA1984.DAT *	/geneseqp-emb1/AA1985.DAT *	/geneseqp-emb1/AA1986.DAT *	/geneseqp-emb1/AA1987.DAT *	/geneseqp-emb1/AA1988.DAT *	/geneseqp-emb1/AA1989.DAT *	/geneseqp-emb1/AA1990.DAT *	/geneseqp-emb1/AA1991.DAT *	/geneseqp-emb1/AA1992.DAT *	/geneseqp-emb1/AA1993.DAT *	/geneseqp-emb1/AA1994.DAT *	/geneseqp-emb1/AA1995.DAT *	/geneseqp-emb1/AA1996.DAT *	/geneseqp-emb1/AA1997.DAT *	/geneseqp-emb1/AA1998.DAT *	/geneseqp-emb1/AA1999.DAT *	/geneseqp-emb1/AA2000.DAT *	/geneseqp-emb1/AA2001.DAT *	/geneseqp-emb1/AA2002.DAT *
-----------------------------	-----------------------------	-----------------------------	-----------------------------	-----------------------------	-----------------------------	-----------------------------	-----------------------------	-----------------------------	-----------------------------	-----------------------------	-----------------------------	-----------------------------	-----------------------------	-----------------------------	-----------------------------	-----------------------------	-----------------------------	-----------------------------	-----------------------------	-----------------------------	-----------------------------	-----------------------------

redicted by chance to have a score of the result being printed, al score distribution.

RES

Description
<i>Moraxella catarrhina</i>
<i>Moraxella catarrhina</i>
<i>Haemophilus somnus</i>
<i>Neisseria meningitidis</i>
<i>Neisseria gonorrhoeae</i>
<i>Neisseria meningitidis</i>
<i>Listeria monocytogenes</i>
<i>Listeria</i> p60 protease
<i>Lactococcus lactis</i>
<i>Listeria monocytogenes</i>

Listeria monocytogenes
H. influenzae B45822
H. influenzae B45822
H. influenzae B45822
Lysin, lactococcus
Micromonospora can
Lactococcus lactis
H. influenzae B45822
 HEV-US2 ORF1 protease
 Beta-lytic protease
Chromobacter lytic
C. glutamicum prote
B. subtilis Y081 r
Enterococcus hirae
 Human protein sequ
Streptococcus pyogen
Streptococcus pyogen
 Truncated cellulase
 Active cellulase I
Lyisin carboxy termi
 Human clone vD22-
 Las17 protein. Sals
 Amino acid sequenc
 Human MAG1 polyipep
 Human Nogo protein
 Human NOG0-A prote
 Novel human secret
 Swine HEV ORF 1 pro
Streptococcus pyogen
Streptococcus poly
Streptococcus pyogen
C. glutamicum prote
Drosophila melanog
Drosophila melanog
Drosophila melanog
 Rat nerve growth f

ALIGNMENTS

RESULT 1

ID AAB20106 standard; Protein; 322 AA

AC AAB20106;

DT 23-APR-2001 (first entry)

DE Moraxella catarrhalis BASB10 protein

KW BASB10; infection; otitis media; pneumonia; therapy; diagnosis;
KW antibacterial; antimicrobial; vaccine.

OS Moraxella catarrhalis

FH	Key	Location/Qualifiers
100	100	100
101	101	101
102	102	102
103	103	103
104	104	104
105	105	105
106	106	106
107	107	107
108	108	108
109	109	109
110	110	110
111	111	111
112	112	112
113	113	113
114	114	114
115	115	115
116	116	116
117	117	117
118	118	118
119	119	119
120	120	120
121	121	121
122	122	122
123	123	123
124	124	124
125	125	125
126	126	126
127	127	127
128	128	128
129	129	129
130	130	130
131	131	131
132	132	132
133	133	133
134	134	134
135	135	135
136	136	136
137	137	137
138	138	138
139	139	139
140	140	140
141	141	141
142	142	142
143	143	143
144	144	144
145	145	145
146	146	146
147	147	147
148	148	148
149	149	149
150	150	150
151	151	151
152	152	152
153	153	153
154	154	154
155	155	155
156	156	156
157	157	157
158	158	158
159	159	159
160	160	160
161	161	161
162	162	162
163	163	163
164	164	164
165	165	165
166	166	166
167	167	167
168	168	168
169	169	169
170	170	170
171	171	171
172	172	172
173	173	173
174	174	174
175	175	175
176	176	176
177	177	177
178	178	178
179	179	179
180	180	180
181	181	181
182	182	182
183	183	183
184	184	184
185	185	185
186	186	186
187	187	187
188	188	188
189	189	189
190	190	190
191	191	191
192	192	192
193	193	193
194	194	194
195	195	195
196	196	196
197	197	197
198	198	198
199	199	199

FT /note= "Gly in translation of BASB110 PCR product"

PN WO200100838-A1

PD 04-JAN-2001.

PF 23-JUN-2000; 2000WO-EP05854.

PR 25-JUN-1999; 99GB-0015031.

PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS

PI Thonnard J;

DR WPI; 2001-112459/12

XX Novel BASH10 polypeptides of Moraxella catarrhalis, useful as a
 PT vaccine for treating Moraxella catarrhalis infections -
 PS Claim 1; Page 82-83; 88pp; English.

XX The present sequence is that of BASH10 protein from Moraxella
 CC catarrhalis strain Mc2931 (ATCC 43617), a causative agent of otitis
 CC media in children and pneumonia in adults. The sequence is a
 CC translation of the BASH10 gene coding region (see AAF30047). It
 CC differs by 1 residue from the predicted polypeptide sequence (see
 CC AAB20107) of a BASH10 PCR product, having Ser rather than Gly at
 CC position 104. The invention provides BASH10 polypeptides, and at
 CC polynucleotides encoding them, as well as expression vectors, host
 CC cells and methods for producing BASH10 polypeptides using
 CC recombinant methods. Also claimed is a vaccine composition
 CC comprising a BASH10 polypeptide, an immunogenic fragment of a
 CC acid sequence identity to BASH10, or comprising a polynucleotide
 CC encoding such a polypeptide. A claimed method of diagnosing a
 CC Moraxella infection involves identifying a BASH10 polypeptide or
 CC antibody. A claimed therapeutic composition useful in treating
 CC humans with M. catarrhalis infection comprises at least 1 antibody
 CC directed against a BASH10 polypeptide. BASH10 polypeptides also
 CC have utility in raising specific antibodies, and in screening for
 CC antibacterial drugs.

Sequence 322 AA;

Query Match 100.0%; Score 1670; DB 22; Length 322;
 Best Local Similarity 100.0%; Pred. No. 5.2e-134;
 Matches 322; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTVTIAINSONOKPIRGLIRGVITTTTILAGCAASKPTVSTSGSHRTSGGGLAIGS 60
 DB 1 MTVTIAINSONOKPIRGLIRGVITTTTILAGCAASKPTVSTSGSHRTSGGGLAIGS 60
 QY 61 QVITDSQGVNRYOVKQDPTVSKIAQRYGLNWRIGIHNLSYITTYGWLTLMSGDL 120
 DB 61 QVITDSQGVNRYOVKQDPTVSKIAQRYGLNWRIGIHNLSYITTYGWLTLMSGDL 120
 QY 121 KVERISISSGVNTATTPSPVAVOSSRPVQOHPAVOKPTPPVYVYVQOQAP 180
 DB 121 KVERISISSGVNTATTPSPVAVOSSRPVQOHPAVOKPTPPVYVYVQOQAP 180
 QY 181 VAPPTVEAPFATGSSGMOPRIPVAGTNPVRRFGTATVAGSTVSNMGMFSGRGGDLIN 240
 DB 181 VAPPTVEAPFATGSSGMOPRIPVAGTNPVRRFGTATVAGSTVSNMGMFSGRGGDLIN 240
 QY 241 ASNAGTVIOADHNMDDASIVIOHTNGFVSSYIHKDAQVKTGDPVTRGQRIASMKQPSG 300
 DB 241 ASNAGTVIOADHNMDDASIVIOHTNGFVSSYIHKDAQVKTGDPVTRGQRIASMKQPSG 300
 QY 301 AALFEFRISNGVYVDPPLTVLK 322
 DB 301 AALFEFRISNGVYVDPPLTVLK 322

RESULT 2

AAB20107
 I AAB20107 standard; Protein; 322 AA.

XX AAB20107;
 AC
 XX 23-APR-2001 (first entry)
 DE Moraxella catarrhalis BASH10 protein.
 XX BASH10; infection; otitis media; pneumonia; therapy; diagnosis;
 KM antibacterial; antimicrobial; vaccine.
 CS Moraxella catarrhalis.
 XX

FF Key Location/Qualifiers
 FT MISC-difference 104 /note= "Ser in translation of BASH10 gene"
 PS WO200100838-A1.
 XX 04-JAN-2001.
 XX 23-JUN-2000; 2000MO-BP05854.
 XX 25-JUN-1999; 99GB-0015031.
 PA (SMK) SMITHKLINE BEECHAM BIOLOGICALS.
 PI Thonard J;
 DR WPI 2001.112459/12.
 DR N-PSDB; AAF30047.

PT Novel BASH10 polypeptides of Moraxella catarrhalis, useful as a
 PT vaccine for treating Moraxella catarrhalis infections -
 PS Claim 1; Page 83-84; 88pp; English.

XX The present sequence is that of BASH10 protein from Moraxella
 CC catarrhalis strain Mc2931 (ATCC 43617), a causative agent of otitis
 CC media in children and pneumonia in adults. The sequence is a
 CC translation of the BASH10 gene coding region (see AAF30047). It differs by
 CC 1 residue from the predicted polypeptide sequence (see AAB20106) of
 CC the BASH10 gene product, having Gly rather than Ser at position
 CC 104. The invention provides BASH10 polypeptides, and
 CC polynucleotides encoding them, as well as expression vectors, host
 CC cells and methods for producing BASH10 polypeptides using
 CC recombinant methods. Also claimed is a vaccine composition
 CC comprising a BASH10 polypeptide, an immunogenic fragment of a
 CC acid sequence identity to BASH10, or comprising a polynucleotide
 CC encoding such a polypeptide. A claimed method of diagnosing a
 CC Moraxella infection involves identifying a BASH10 polypeptide or
 CC antibody. A claimed therapeutic composition useful in treating
 CC humans with M. catarrhalis infection comprises at least 1 antibody
 CC directed against a BASH10 polypeptide. BASH10 polypeptides also
 CC have utility in raising specific antibodies, and in screening for
 CC antibacterial drugs.

Sequence 322 AA;

Query Match 99.8%; Score 1666; DB 22; Length 322;
 Best Local Similarity 99.7%; Pred. No. 1.1e-133;
 Matches 321; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MTVTIAINSONOKPIRGLIRGVITTTTILAGCAASKPTVSTSGSHRTSGGGLAIGS 60
 DB 1 MTVTIAINSONOKPIRGLIRGVITTTTILAGCAASKPTVSTSGSHRTSGGGLAIGS 60
 QY 61 QVITDSQGVNRYOVKQDPTVSKIAQRYGLNWRIGIHNLSYITTYGWLTLMSGDL 120
 DB 61 QVITDSQGVNRYOVKQDPTVSKIAQRYGLNWRIGIHNLSYITTYGWLTLMSGDL 120
 QY 121 KVERISISSGVNTATTPSPVAVOSSRPVQOHPAVOKPTPPVYVYVQOQAP 180
 DB 121 KVERISISSGVNTATTPSPVAVOSSRPVQOHPAVOKPTPPVYVYVQOQAP 180
 QY 181 VAPPTVEAPFATGSSGMOPRIPVAGTNPVRRFGTATVAGSTVSNMGMFSGRGGDLIN 240
 DB 181 VAPPTVEAPFATGSSGMOPRIPVAGTNPVRRFGTATVAGSTVSNMGMFSGRGGDLIN 240
 QY 241 ASNAGTVIOADHNMDDASIVIOHTNGFVSSYIHKDAQVKTGDPVTRGQRIASMKQPSG 300
 DB 241 ASNAGTVIOADHNMDDASIVIOHTNGFVSSYIHKDAQVKTGDPVTRGQRIASMKQPSG 300
 QY 301 AALFEFRISNGVYVDPPLTVLK 322
 DB 301 AALFEFRISNGVYVDPPLTVLK 322

```

DB          301 AALFEFRISRNQGVYVDPLTVLK 322

RESULT 3
ID AAB21225
XX AAB21225 standard; Protein; 345 AA.
AC AAB21225;
DT 12-JAN-2001 (first entry)
XX
XX Haemophilus somnus lipB.
KW Haemophilus somnus; lipoprotein B; lipB; transferrin-binding protein;
KW lbp; antibacterial; vaccine; infection;
KW chromoembolic meningoencephalitis; ITMEM; myocarditis; septicæmia;
XX arthritis; pneumonia.
XX Haemophilus somnus.
XX
XX WO200053765-A1.
XX
XX 14-SEP-2000.
XX
XX 10-MAR-2000; 2000WO-CA00244.
XX
XX 10-MAR-1999; 99US-0267749.
XX 24-SEP-1999; 99US-0405728.
XX
XX (UNSA-) UNIV SASKATCHEWAN.
XX
XX Potter AA, Rioux C, Schryvers AB,
XX WPI; 2000-565601/52.
XX DR N-PSDB; AAA93349.
XX
XX Nucleic acids encoding Haemophilus somnus transferrin binding proteins
XX useful for vaccinating against and diagnosing H. somnus infections e.g.
XX myocarditis and pneumonia in cattle -
XX
XX Disclosure; Fig 11; 67pp; English.
XX
XX
XX The present sequence is the Haemophilus somnus lipoprotein lipB.
XX Recombinant vaccines containing H. somnus lipoprotein lipB, lipB
XX and lppC have been described. H. somnus transferrin-binding proteins Tbp1
XX and Tbp2 have also been found to be effective in vaccines against
XX H. somnus infection. Clones expressing Tbp epitopes were identified by
XX screening a genomic expression library of H. somnus strain H825 in
XX E. coli with polyclonal antiserum raised against affinity-purified Tbp1
XX and Tbp2 of H. somnus. The genes coding for Tbp1 and Tbp2 were obtained
XX by inverse PCR. The nucleotide sequence encoding Tbp1 and Tbp2 may be
XX used for the recombinant production of Tbp1 and Tbp2, which may then be
XX used to manufacture vaccine compositions for immunising against H. somnus
XX infections. The antibodies raised against the transferrin binding
XX proteins may also be used diagnostically to identify the presence of
XX H. somnus infections. H. somnus is a pathological Gram-negative bacteria
XX that causes a number of diseases in cattle such as chromoembolic
XX meningoencephalitis (ITMEM), myocarditis, septicæmia, arthritis and
XX pneumonia.
XX
XX Sequence 345 AA;

Query Match 12.5%; Score 209; DB 21; Length 345;
Best Local Similarity 24.8%; Pred. No. 1.4e-09;
Matches 63; Conservative 49; Mismatches 106; Indels 36; Gaps 6

CY 73 YQYKGGDTVSKAQRVYGLNWRKIEGHINNINSYTYITGQWLTMSG---DLKVERRSIS 129
DB 120 YKVRGDTMTFLIAYISGMDIKELATLNMSSEPHYSIGCVLKIANNIPSNMIPQTIWE 179
CY 130 GVNTAHTSPVAVQSSRPVQGNHAPVQKTPPVVVVKKRTPRPVVQGNAPAPRPTEAP 169
DB 180 SEVITQNTVNE-TWNAANKPTNEQMKFVAPITHSIMPTLNK--TPPATSNIAIMTWP----- 229

```

Qy	190	FATGSSGVMGRYRPGAINPVRREGTATVAGSTYTSNGMMRSGRDGDLINNSNACTVIG	243
Db	230	-----TNGKTIHQGFSSADG-----NKGIDISGRGAQVAAAAGRVVY	268
Qy	250	ADHNMDGAS--IVIQHTNGFVSSSYTHIKDAQYKGTDTVTRGQRIASMKNPQSGAALFEFR	307
Db	269	AGDALRGVGNLIITKHNDSYLSAVAHNNSILVKKQGFVAKQGIAMKSSGNTITKLHRE	328
Qy	308	ISRNGVYVDPLTVL	321
Db	329	IRYKQGSYDPMRKYL	342
RESULT 4			
AA74313	ID	AA74313 standard; Protein; 414 AA.	
AA74313;	AC		
AA74313;	DT	21-MAR-2000 (first entry)	
XX	DE	Neisseria meningitidis ORF 025 protein sequence SEQ ID NO:112-1.	
XX	XX		
XX	KW	Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;	
XX	KW	antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;	
XX	XX	antibacterial; gene therapy.	
XX	OS	Neisseria meningitidis.	
XX	PN	W0957280-A2.	
XX	PD	11-NOV-1999.	
XX	PF	30-APR-1999; 99WO-US09346.	
XX	PR	01-MAY-1998; 98US-0083758.	
XX	PR	31-JUL-1998; 98US-0094869.	
XX	PR	02-SEP-1998; 98US-0098994.	
XX	PR	02-SEP-1998; 98US-0099062.	
XX	PR	09-OCT-1998; 98US-0103749.	
XX	PR	09-OCT-1998; 98US-0103794.	
XX	PR	09-OCT-1998; 98US-0103796.	
XX	PR	25-FEB-1999; 99US-0121528.	
XX	PA	(CHIR) CHIRON CORP.	
XX	PA	(GENO-) INST GENOMIC RES.	
XX	PI	Fraser C, Galeotti C, Grandi G, Hickey E, Massignani V, Mora M;	
XX	PI	Petersen J, Piza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;	
XX	PI	Teteltn H, Venter JC;	
XX	DR	WPI; 2000-062150/05.	
XX	DR	N-PSDB; AAZ53075.	
XX	PT	Novel Neisserial polypeptides predicted to be useful antigens for	
XX	PT	vaccines and diagnostics -	
XX	PS	Claim 2; Page 207; 1453pp; English.	
XX	XX	AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941	
XX	XX	represent novel Neisseria meningitis and N. gonorrhoeae polynucleotides	
XX	XX	and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ5473 represent	
XX	XX	PCR primers used in the exemplification of the present invention. The	
XX	XX	polypeptides, the polynucleotides, antibodies and compositions of	
XX	XX	the invention can be used as vaccines, as diagnostic reagents, and as	
XX	XX	immunogenic compositions. The polypeptides can be used in the	
XX	XX	manufacture of medicaments for treating or preventing infection due to	
XX	XX	Neisserial bacteria (e.g. meningitis and septicemia), to detect the	
XX	XX	presence of Neisseria bacteria, or to raise antibodies. They may also	
XX	XX	be used to screen for agonists or antagonists, which may themselves	
XX	XX	have use as antibacterial agents. The polynucleotides of the invention	
XX	XX	may also be used in gene therapy protocols.	

XX Sequence 414 AA;
SQ Query Match 12.2%; Score 203; DB 21; Length 414;
Best Local Similarity 27.3%; Pred. No. 5,8e-09;
Matches 71; Conservative 42; Mismatches 97; Indels 50; Gaps 11;
QY 73 YVKGQDPTVSKIAQRGLNMRREIGHNINNSYTYTQWLTMSGDLKYRERSISSGVN 132
DB 184 HTIVRGDPTVYNSIKRKHISODDFRANMGMTDN-TLSIQO-----IVKVPAGYAA--- 232
QY 133 THTSPVAVOSSRPVQOHPAVQKPTPPVVVKKPTPTPPVVOQAPAPVAP----- 183
DB 233 ----PKAAVVK-SRPVVP--AAVQTPV-----KPAQPPVQSAPOPAPAPAAENKAVPA 278
QY 184 PVTAPFA----TGSSGVWQFRYPVGAATNPVVRREGTATVAGSTVTSNGMWFSGRGGDLI 239
DB 279 PAPQSPAAISPGSTRSVGCIWQRP--TQKVVADFG-----NNKGVDIAGNAQPV 328
QY 240 NNSMGCTVIOADHNDGAS--IVIOHTNGFVSSYTHIKDAQVKTGDTVRGQRIASMKQ 237
DB 329 LPAADGKVYVAGSGLRGYNLVITIQHNSFLTAIGHNCKLLVGGQGVKRGQVAAVMGNT 388
QY 298 PSGALFEFRISRNQVYVDP 317
DB 389 EASRTQLHFEVRNGKPVNP 408
RESULT 5
AAI74311
ID AAY74311 standard; Protein; 337 AA.
XX AAY74311;
XX 21-MAR-2000 (first entry)
DE Neisseria gonorrhoeae ORF 025 protein sequence SEQ ID NO:118.
XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KM antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
KW antibacterial; gene therapy.
OS Neisseria gonorrhoeae.
XX PN WO9957280-A2.
XX 11-NOV-1999.
XX 30-APR-1999; 99WO-US09346.
XX 01-MAY-1998; 98US-0083758.
XX 31-JUL-1998; 98US-0094869.
XX 02-SEP-1998; 98US-0098994.
XX 02-SEP-1998; 98US-0099062.
XX 09-OCT-1998; 98US-0103749.
XX 09-OCT-1998; 98US-0103794.
XX 09-OCT-1998; 98US-0103796.
XX 25-FEB-1999; 99US-0121528.
XX (CHIR) CHIRON CORP.
XX (GENO-) INST GENOMIC RES.
XX PI Frazer C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M,
XX PI Petersen J, Piazza M, Rappuoli R, Ratti G, Scalato E, Scarselli M,
XX PI Tettein H, Venter JC;
XX WPI, 2000-062150/05.
XX N-PSDB; AAZ53073.
XX Novel Neisserial polypeptides predicted to be useful antigens for
XX vaccines and diagnostics
XX Claim 2; Page 206; 1453pp; English.

XX AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
CC and polypeptides; AAZ54537 to AAZ54576 and AAZ54616 to AAZ5473 represent
CC PCR primers used in the exemplification of the present invention. The
CC polypeptides, the polynucleotides, antibodies and compositions of
CC the invention can be used as vaccines, as diagnostic reagents, and as
CC immunogenic compositions. The polypeptides can be used in the
CC manufacture of medicaments for treating or preventing infection due to
CC Neisseria bacteria (e.g. meningitis and septicemia), to detect the
CC presence of Neisseria bacteria, or to raise antibodies. They may also
CC be used to screen for agonists or antagonists, which may themselves
CC have use as antibacterial agents. The polynucleotides of the invention
CC may also be used in gene therapy protocols.
XX SQ Sequence 337 AA;
Query Match 12.0%; Score 200; DB 21; Length 337;
Best Local Similarity 26.2%; Pred. No. 8e-09;
Matches 66; Conservative 45; Mismatches 109; Indels 32; Gaps 9;
QY 73 YVKGQDPTVSKIAQRGLNMRREIGHNINNSYTYTQWLTMSGDLKYRERSISSGVN 132
DB 105 HTIVRGDPTVYNSIKRKHISODDFRANMGMTDN-TLSIQO-----IVKVPAGYAA--- 156
QY 133 THTSPVAVOSSRPVQOHPAVQKPTPPVVVKKPTPTPPVVOQ--PAPVAPVTEA- 188
DB 157 AAVESRPVAPVAAQTPVK--PAAQPP-----VQSAQOPAPAPAAENKAVPAPAPQSPAA 209
QY 189 -PFATGSSGVWQFRYPVGAATNPVVRREGTATVAGSTVTSNGMWFSGRGGDLINASNAQTV 247
DB 210 SPSTGRSVGCIWQRP--TQKVVADFGG-----NNKGVDIAGNAQPVIAADGKV 259
QY 248 IOADHNDGAS--IVIOHTNGFVSSYTHIKDAQVKTGDTVRGQRIASMKQPSGAALE 305
DB 260 VYAGSGLRGYNLVITIQHNSFLTAIGHNCKLLVGGQGVKRGQVAAVMGNTASRTQLH 319
QY 306 FRISRNQVYVDP 317
DB 320 FEVRNGKPVNP 331
RESULT 6
AAI74312
ID AAY74312 standard; Protein; 351 AA.
XX AAY74312;
XX 21-MAR-2000 (first entry)
DE Neisseria meningitidis ORF 025 protein sequence SEQ ID NO:110-1.
XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KM antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
KW antibacterial; gene therapy.
OS Neisseria meningitidis.
XX PN WO9957280-A2.
XX 11-NOV-1999.
XX 30-APR-1999; 99WO-US09346.
XX 01-MAY-1998; 98US-0083758.
XX 31-JUL-1998; 98US-0094869.
XX 02-SEP-1998; 98US-0098994.
XX 02-SEP-1998; 98US-0099062.
XX 09-OCT-1998; 98US-0103749.
XX 09-OCT-1998; 98US-0103794.
XX 09-OCT-1998; 98US-0103796.
XX 25-FEB-1999; 99US-0121528.

PA (CHIR) CHIRON CORP.
 PA (GENO-) INST GENOMIC RES.
 XX
 PI Praeger C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
 PI Petersen J, Pizzi M, Rappoli R, Ratti G, Scalato E, Scarselli M,
 PI Tettein J, Venter JC;
 XX
 DR WPI; 2000-062150/05.
 DR N-PSDB; AA253074.
 PT Novel Neisserial polypeptides predicted to be useful antigens for
 PT vaccines and diagnostics -
 XX
 PS Claim 2; Page 206; 1453pp; English.
 CC AA253015 to AA254536, AA254577 to AA254615, and AA274253 to AA275941
 CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
 CC and polypeptides. AA254537 to AA254576 and AA254616 to AA25473 represent
 CC PCR primers used in the exemplification of the present invention. The
 CC polypeptides, the polynucleotides, antibodies and compositions of
 CC the invention can be used as vaccines, as diagnostic reagents, and as
 CC immunogenic compositions. The polypeptides can be used in the
 CC manufacture of medicaments for treating or preventing infection due to
 CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the
 CC presence of Neisseria bacteria, or to raise antibodies. They may also
 CC be used to screen for agonists or antagonists, which may themselves
 CC have use as antibacterial agents. The polynucleotides of the invention
 CC may also be used in gene therapy protocols.
 CC
 XX Sequence 351 AA;
 SQ
 Query Match 11.9%; Score 198; DB 21; Length 351;
 Best local similarity 27.2%; Pred. No. 1.3e-08;
 Matches 72; Conservative 38; Mismatches 95; Indels 60; Gaps 12;
 QY 73 YQVKGQDVSKIAQRYGLNMRREIGHININSSYTYTGQWLTMSGDLKVRERSISG-- 130
 DB 121 HTIVGDTVYNISKRY-----HISQDD-----FRAMNG---MTDNMLSIGOI 159
 QY 131 --VNTAHTPSP-VAVQSSRPVQOHAPVQKPPVPPVVKKPPPTPPVQGPAPVAP---- 183
 DB 160 VKVKPAGVAAAPKTAVERSPAPV--AAVQTPV-----KPAQPPVQAPQAPAPAAEN 210
 QY 184 -----PVTAPPA-----TGSSGVMQPRYPVGAATNPVRRFGATVAGSTVTSNGMVFSGR 234
 DB 211 KAVPAPAPQSPAPASPSGSTRSVGIVWQRP--TQGVVADFG-----NKKGVDIAGN 260
 QY 235 DGDLLINASNAQTVIQADHNMDSAS--IVIQHTNGFVSSYTHIKDAQVKTGDTVRTGORIA 292
 DB 261 AGQPVLAADGKVVVAGSGLRGVGNLVIIOHNSFLTAIGHNQKLLVGGQGVKRGQOVA 320
 QY 293 SKKNQPSGAALEFFRISRNGVYVP 317
 DB 321 LMGNTDASRTQLHFEVRONGKEVPNP 345
 DB
 RESULT 7
 ABB48721
 ID ABB48721 standard; Protein; 436 AA.
 XX
 AC ABB48721;
 DT 05-FEB-2002 (first entry)
 DE Listeria monocytogenes protein #1425.
 KW Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
 KW vitamin B12; bacterial infection; disease.
 OS Listeria monocytogenes.
 XX
 XX WO200177335-A2.
 XX

PD 18-OCT-2001.
 XX
 PF 11-APR-2001; 2001WO-FR01118.
 XX
 PR 11-APR-2000; 2000FR-0004629.
 XX
 PA (INSP) INST PASTEUR.
 XX
 PI Buchrieser C, Frangeul L, Couve E, Rusniok C, Fsihi H, Dehoux P;
 PI Dussauget O, Cherouni F, Nedjari H, Glaser P, Kunst F, Cossart P;
 PI Daniels U, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA,
 PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;
 PI Chakraborty T, Domann B, Hain T, Berche P, Charbit A, Durant L;
 PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;
 PI Madueno E, De Pablos B, Wehlund J, Kaerst U, Entian K, Hauf J;
 PI Rose M, Voss H;
 XX
 DR WPI; 2002-010914/01.
 PT Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
 PT and prevention of Listeria and related bacterial infections, and
 PT related polypeptides -
 XX
 PS Claim 6; SEQ ID No 1426; 192pp; French.
 CC
 CC The present invention relates to the genome sequence of Listeria
 CC monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of
 CC it are useful for selecting probes and primers for detecting genes in L.
 CC monocytogenes and related organisms, and for studying genetic
 CC polymorphisms and other genomes. The present sequence is a protein
 CC encoded by the genome sequence of the present invention. Proteins
 CC expressed from the genome sequence are useful for raising specific
 CC antibodies, identification of L. monocytogenes and related organisms, and
 CC for biosynthesis and biodegradation, especially biosynthesis of vitamin
 CC B12. The genome sequence and proteins encoded by it are also useful for
 CC selecting compounds that regulate gene expression and cell replication
 CC and modulate L. monocytogenes-related diseases. In addition, the genome
 CC sequence and proteins encoded by it are useful in pharmaceutical and
 CC vaccines compositions for the treatment or prevention of infections by L.
 CC monocytogenes and related organisms.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX Sequence 436 AA;
 SQ
 Query Match 9.5%; Score 158.5; DB 23; Length 436;
 Best local similarity 24.8%; Pred. No. 3.8e-05;
 Matches 66; Conservative 33; Mismatches 100; Indels 67; Gaps 11;
 QY 76 KQDPTVSKIAQRYGLNMRREIGHININSSYTYTGQWLTMSGDLKVRERSISGVNTAH 135
 DB 210 EKNDLVMAALANKKDLT-----KSEQTLASE-----QGALTDEKKLASINIGEK 254
 QY 136 TPSPVAVQSSRPVQOHAPVQKPPVPPVVKKPPPTPPVQGPAPVAPVTEAFATGSS 195
 DB 255 AKQERATKAABEKMQEAAASAKSAAYK-----QPSSSSEATE-----TVSS 300
 QY 196 GVMQPRYPVGA-----TNPVVRP-----GTATVAGSTVTSNGMVFSGRGGDLINAS 242
 DB 301 GGGQFIRKPSAGILTSGFSERTNPVTGKXESHKGODIAGGTVT-----VSA 347
 QY 243 NAGTVIQADHNMDSAS-----IVIQHTNGFVSSYTHIKDAQV--TGDVTRTGORIAS 293
 DB 348 ASGTVVPSGPAASGSGGCGYGVVKKIDGNSFQTLGHMRAGSLKVVTTGQGVSGQSPIGI 407
 QY 294 M--KNQPSGAALEFFRISRNGVYVP 317
 DB 408 MGTGQSTGCHL-HPEIHKNGIPIVP 432
 DB
 RESULT 8
 AAR45178

ID AAR45178 standard; Protein; 478 AA.
AC AAR45178;
XX
DT 16-JUN-1994 (first entry)
XX
DE Listeria p60 protein.
XX
KW Immunogenic polypeptide; antibodies; immunoassay; conjugate.
OS Listeria monocytogenes.
XX
PN DE4318450-A.
XX
PD 16-DEC-1993.
XX
PP 03-JUN-1993; 93DE-4318450.
XX
PR 11-JUN-1992; 92DE-4219111.
XX 23-NOV-1992; 92DE-4239567.
XX
PA (MERIE) MERCK PATENT GMBH.
XX
PI Schubert P, Neumann S, Pawelzik M, Linkweiller W, Burger C;
XX Hofmann G, Buber A, Goebel W, Koehler S;
DR WPI; 1993-406956/51.
XX
PT New primers for PCR detection of Listeria - including individual
XX species, also new peptide(s) for raising antibodies for
XX immunochemical detection
PS Disclosure; Fig 4; 19pp; German.
XX
CC The sequence is that of the Listeria p60 protein. Antibodies
CC generated against the protein can be used in the detection
CC of Listeria by immunoassay (partic. ELISA). The detection method
CC allows determination of individual Listeria species, esp.
XX L. monocytogenes.
SQ Sequence 478 AA;
Query Match 9.1%; Score 151.5; DB 14; Length 478;
Best Local Similarity 21.3%; Pred. No. 0.00017; Mismatches 117; Indels 143; Gaps 15;
Matches 83; Conservative 46; Mismatches 117; Indels 143; Gaps 15;
QY 25 ITTCLAGCASKPTYNSTSGSGSHRTSGGGLA--IGSQVITD----- 65
DB 102 IITSIKG--TKVIVETTESNGWKITVNDGKTGFVNGKYLTDKAVSTPAVTOEVKKT 159
QY 66 --SQGVP-----NRVQVKGADTVSKIAQRYG 89
DB 160 TTQQAAPVAETKTEVKQTQATTPAKVAETETVIDONATTAAVKSGLTIALSVKYG 219
QY 90 LNMREIGHINNLSSTIYTGOMLTMSGDLKVRERSISSGVNTAHTPSVVAOVSRRPV 149
DB 220 VSVQDIMSNNLSS--SIYVGOKLAI-----KOTANATPKAEVKT----- 259
QY 150 QOHFAVQKTPPVV-----VKKPTPPVVOQAPVAPVTE-----APFATG 193
DB 260 -EAPAAEKQAPVKEKNTNTNTATKEKEITQ--OOTAPKAP--TEAAKPAAPSTNT 313
QY 194 SSGVMOFRYPVATPVVRRFGTAVAGSTVTSNGMFGSGRDDLINAAGTAVIOADHN 253
DB 314 NANKTNTNTNTNTNTTPEKNTNTNTNTNTNTNANOGSSNNNSNSASALIAEQKH 373
QY 254 MD-----GASIVIOHTNGF--VSSYIHKDAQVGTGTVRTG 288
DB 374 LKAYSMGANGPTTDCSGYTKVFAAGISLERTISGAQYASTTRISBQAKPDIV--- 430
QY 289 QRIASMKQPSGALFFP--RISNGVTV 315
DB 431 -----FFDYSGSISHVGIIV 445

RESULT 9
ID ABB53578 standard; Protein; 439 AA.
XX ABB53578;
XX
AC ABB53578;
XX
DT 16-MAY-2002 (first entry)
XX
DE Lactococcus lactis protein acma.
XX
KW Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.
OS Lactococcus lactis IL1403.
XX
PN FR2807446-A1.
XX
PD 12-OCT-2001.
XX
PP 11-APR-2000; 2000FR-0004630.
XX
PR 11-APR-2000; 2000FR-0004630.
XX
PA (INRG) INRA INST NAT RECH AGRONOMIQUE.
XX
PI Bojorine A, Sorokine A, Renault P, Ehrlich SD;
XX WPI; 2002-043418/06.
XX
PT New nucleotide sequence useful in the identification of Lactococcus
XX lactis and related species
PS Claim 6; SEQ ID No 280; 2504pp; French.
XX
CC The present invention is related to a Lactococcus lactis nucleotide
CC sequence (ABA90521) and related proteins (ABB53300-ABB55621). The
CC nucleic acid sequence is useful in the detection and/or amplification of
CC related species. The proteins of the invention are useful for the
CC biosynthesis or biodegradation of the invention are useful for the
CC production of yogurt and cheese.
CC Note: The sequence data for this patent is based on equivalent patent
CC WO2001/77334 (published 18-OCT-2001) which is available in electronic
XX format directly from WIFO at ftp.wifo.int/pub/published_pct_sequences.
SQ Sequence 439 AA;
Query Match 8.8%; Score 146.5; DB 23; Length 439;
Best Local Similarity 23.6%; Pred. No. 0.00041; Mismatches 69; Conservative 39; Mismatches 107; Indels 77; Gaps 10;
Matches 69; Conservative 39; Mismatches 107; Indels 77; Gaps 10;
QY 26 TTCIAGCASKPTYNST-----SSGSHRTSGGGLAIGSQVITDQVPR 72
DB 184 TAAITGKATDPRNGALNRIISQYNLTRFDGASAGTSGG--STANTNNNSNTSST 242
QY 73 YOVKGDTVSKIAQRYGLMREIGHINNLSSTIYTGOMLTMSGDLKVRERSISSGVN 132
DB 243 YTVASGDTLNGISQKGISVAQISANNLKT--VYIGOKLVLTSSSSNTNTSSSGN 301
QY 133 TAHTPSVVAOVSRRPVVQOHAPVQKPTPPVVKPTPPVVOQAPVAPVPT---EA 188
DB 302 SAGTTTFTT-----SVTPAPASQITIKYSGDTLMGLSV 336
QY 189 PFATGSSGMORRYPVGATNPVRRFGTAVAGSTVTSNGMFGSGRDDLINAAGTAVI 248
DB 337 KYVTITAGLQKSMNNLNSDTIFIGNLIVSOSAGSSSSSTG-----SSAST-- 382
QY 249 QADHNNDDGASIVIOHTNGFVSSYIHKDAQVGTGTV-----RTGQRTASMK 295
DB 383 SGTSMNSAAS-----NTSIH--KVVKGDTLMGLSQSGSPIASIK 420

RESULT 10
 ABB48659 standard; Protein; 482 AA.
 ID ABB48659 standard; Protein; 482 AA.
 AC ABB48659;
 XX ABB48659;
 XX ABB48659;
 DT 05-FEB-2002 (first entry)
 XX Listeria monocytogenes protein #1363.
 DE Listeria monocytogenes protein #1363.
 XX Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
 KW vitamin B12; bacterial infection; disease.
 XX Listeria monocytogenes.
 OS Listeria monocytogenes.
 FN WO200177335-A2.
 XX WO200177335-A2.
 PD 18-OCT-2001.
 XX 18-OCT-2001.
 PF 11-APR-2001; 2001WO-FR01118.
 XX 11-APR-2001; 2001WO-FR01118.
 PR 11-APR-2000; 2000FR-0004629.
 XX 11-APR-2000; 2000FR-0004629.
 PA (INSP) INST PASTEUR.
 XX (INSP) INST PASTEUR.
 PI Buchrieser C, Frangoul L, Couve E, Rusniok C, Fathi H, Deboux P,
 PI Duesurget O, Chetoui F, Nedjati H, Glaeser P, Kunst F, Coesart P,
 PI Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA,
 PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A,
 PI Chakraborty T, Doman E, Hain T, Berche P, Charbit A, Durant L,
 PI Perez-Diaz J, Baguerio F, Garcia Del Portillo F, Gomez-Lopez N,
 PI Madueno B, De Pablo B, Wehlend J, Kaerst U, Entian K, Hauf J,
 PI Rose M, Voss H;
 XX WPI; 2002-010914/01.
 DR WPI; 2002-010914/01.
 PT Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
 PT and prevention of Listeria and related bacterial infections, and
 PT related polypeptides -
 XX related polypeptides -
 PS Claim 6; SEQ ID No 1364; 192pp; French.
 XX Claim 6; SEQ ID No 1364; 192pp; French.
 XX The present invention relates to the genome sequence of Listeria
 CC monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of
 CC it are useful for selecting probes and primers for detecting genes in L.
 CC monocytogenes and related organisms, and for studying genetic
 CC polymorphisms and other genomes. The present sequence is a protein
 CC encoded by the genome sequence of the present invention. Proteins
 CC expressed from the genome sequence are useful for raising specific
 CC antibodies, identification of L. monocytogenes and related organisms, and
 CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin
 CC B12. The genome sequence and proteins encoded by it are also useful for
 CC selecting compounds that regulate gene expression and cell replication
 CC and modulate L. monocytogenes-related diseases. In addition, the genome
 CC sequence and proteins encoded by it are useful in pharmaceutical and
 CC vaccine compositions for the treatment or prevention of infections by L.
 CC monocytogenes and related organisms.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIDO
 CC at ftp.wido.int/pub/published_pct_sequences.
 XX at ftp.wido.int/pub/published_pct_sequences.
 XX Sequence 482 AA;
 SQ Sequence 482 AA;
 Query Match 8.8%; Score 146.5; DB 23; Length 482;
 Best Local Similarity 20.6%; Pred. No. 0.00046;
 Matches 81; Conservative 46; Mismatches 121; Indels 145; Gaps 14;
 QY 25 ITTCLACASKPTNSTSGSGSHRTSSGGLA--IGSQVITD----- 65
 DB 100 IITSTKGG--TKVIVETTESNGMHTKITYNDGTGTVNGSKYLLDKAVSTPVAFTQEVKKT 157
 QY 66 --SQGVV-----NRVVGKGDVTSKIAQRVY 89

Db 158 TTQOAPPAETKTEVKQTQATTPAPKVAETKETPVVDQNAHTAVASGDTIMALSVKYG 217
 QY 90 LNMREIGHINNLNSSTYTYTQGLTMSGDLKVAERISISSGVNNAHPSPPAVOSSRPV 149
 Db 218 VSVQDIMSNNLSSS-STYVQOKLAI-----KQANATTPPAEVKT----- 257
 QY 150 QQHFAVCKPTPPV-----VKKPTTPPVQOPAPVAPVTEAPATGSSGVWQ 199
 Db 258 -EAPPAEKQAPVVKENNTNTATTEKETITQ---QQTAPKATPEAKPAPASTTNA 313
 QY 200 FRYPVGA-----TNPVRRFGTAVAGSTVTSNGMFGSRDGLINASNAVITQ 249
 Db 314 NKTNTNTNTNTNTNTNTNTPTSPKNTNTNSNTNTNSNTNANQSSNNNSASATIAE 373
 QY 250 ADHMD-----GASVIOHTNGF-VSSYHIMDAQVTKGDT 284
 Db 374 AQKHGKAVSWGNGPPTFDGSGYTKYVFAVAGISLPTSGAAYSTTRISEQAKPGDL 433
 QY 285 VRTGQRIASMKQPSGALPEF--RISRNGVYV 315
 Db 434 V-----FPDYGSGISVHGIV 449
 RESULT 11
 AAR73913
 ID AAR73913 standard; protein; 484 AA.
 AC AAR73913;
 XX AAR73913;
 DT 05-DEC-1995 (first entry)
 XX 05-DEC-1995 (first entry)
 DE Listeria monocytogenes protein p60 precursor.
 XX Listeria monocytogenes protein p60 precursor.
 KW Listeria monocytogenes; protein p60 precursor; vaccine;
 KW meningitis related homologous antigenic sequence; MRHAS; RV-1;
 KW immunosassay; diagnosis; treatment; prophylactic; bacterial;
 KW viral.
 XX Listeria monocytogenes.
 OS Listeria monocytogenes.
 FN WO9509232-A.
 XX WO9509232-A.
 PD 06-APR-1995.
 XX 06-APR-1995.
 PF 28-SEP-1994; 94WO-CA00516.
 XX 28-SEP-1994; 94WO-CA00516.
 PR 28-SEP-1993; 93US-0127499.
 XX 28-SEP-1993; 93US-0127499.
 PA (SHAR/) SHARMA L R.
 PA (VALS/) VAN ALSTYNE D.
 PI Sharma LR, Van Alstyne D;
 DR WPI; 1995-147431/19.
 PT New peptide(s) and corresp. antibodies for the treatment of
 PT meningitis - the peptide(s) corresp. to homologous antigenic
 PT sites on bacterial and viral agents and on chemokine(s), used for
 PT detecting and preventing meningitis
 XX detecting and preventing meningitis
 XX Claim 47; Fig 7/10; 98pp; English.
 CC AAR73913 is the Listeria monocytogenes protein p60 precursor. It
 CC contains the meningitis related antigenic sequences (MRHAS) claimed
 CC in AAR73891-R73894 and AAR73903-R73906, which are recognised by a
 CC monoclonal antibody from the hybridoma Rubella virus (RV)-1. The
 CC claimed MRHAS peptides may be used in immunoassays to diagnose the
 CC presence of bacterial and/or viral meningitis agents in a sample,
 CC or in prophylactic and therapeutic meningitis treatments. The
 CC peptides may also be used as vaccines against meningitis.
 CC NB: Identified by matching corresponding MRHAS peptides.

us-10-018-706-2.rag

Sequence 484 AA;

```

Query Match      8.8%; Score 146.5; DB 16; Length 484;
Best Local Similarity 20.6%; Pred. No. 0.00046;
Matches      81; Conservative 46; Mismatches 11

```

useful as components of vaccines for treating bacterial infection such as otitis media, delayed speech learning and inflammation of middle ear

The present invention provides the protein and coding sequences of several versions of the BASH201 protein from non-typable *Haemophilus influenzae*. These can be used in the production of vaccines against *H. influenzae* infection, which can cause otitis media in infants and children, pneumonia in elders, sinusitis, otospecoacal infections, or invasive diseases, chronic otitis media with effusion, or otitis media with effusion in the middle ear, auditive nerve deafness, fluid accumulation in the middle ear, and/or delayed speech development. The present sequence is a version of the BASH201 protein sequence of the invention.

SQ Sequence 410 AA;

Query Match	8.0%;	Score 134;	DB 23;	Length 410;
Best Local Similarity	24.6%;	Pred. No. 0.0043;		
Matches	63;			

	62; conservative	37;	Mismatches III;	Indels	42;	Gaps	10
OY	74 QVKQGDVTSKIAORYSINMPRECHTINNUNSCVMTMTEGNTN						

Db
188 QKQOQALQKAKQEQHS-----LNELMKNIALLDQK-----LNALKANEGALPQET 329

134 AHTEPVAVQSSRPV-QQHPAVQKPTPPVVVKKPTPTPPVQGPAPVAPVTEAPFAT 192

-----SKPYQPTVERQLLN 278

193 GSSGV---MORRYPVGATNPVRRFGTAVAGSTMYMCNCCNCCNC

279 STSLGAKKQYSLPVGSG--ILHTFG--SIAGEFRMKGMVIGASAGTGPVAIAAGRTV 334

```

249 QADH-NMDGASIVIOHTNGFVSSYIHKDAQVKTGDTVRTGTGRISMKN--QPSGALFE 305
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
335 LAGYINGYGVMTVTKRGRDRIYR.....: : : : : : : : : : : : : : : : :

```

306 FRISRNGVDP 317

b
394 FGISRKGTVPVN 405

RESULT 13
AO17656

AA017656 standard; Protein, 410 AA.
AA017656;
AA017656

05-AUG-2002 (first entry)

H influenzae BASB201 #1.

measles; otitis media; pneumonia; sinusitis; nosocomial infection; auditive nerve damage; delayed speech learning; vaccine; antibacterial; auditory; antiinflammatory

Haemophilus influenzae.

Key	Location/Qualifiers
Domain	32..300
	72..

```
Domain
/label= KEO_rich_domain
301.400
/label= peptidase M37-like domain
```

MO200230967-A2.

05-OCT-2001 : 2001WO-FBI1CCC

13-OCT-2000; 2000GB-0025169.

XX (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

XX Thonard J;

XX WPI; 2002-426267/45.

DR N-PSDB; AAL46624.

XX New isolated non-typable Haemophilus influenzae BASB201 polypeptides,
PT useful as components of vaccines for treating bacterial infection such
PT as otitis media, delayed speech learning and inflammation of middle ear
PT

PS Claim 3; Page 87; 90pp; English.

XX The present invention provides the protein and coding sequences of
CC several versions of the BASB201 protein from non-typable Haemophilus
CC influenzae. These can be used in the production of vaccines against H.
CC influenzae infection, which can cause otitis media in infants and
CC children, pneumonia in elders, sinusitis, nosocomial infections, or
CC invasive diseases, chronic otitis media with hearing loss, fluid
CC accumulation in the middle ear, auditive nerve damage, delayed speech
CC learning, infections of the upper respiratory tract and inflammation of
CC the middle ear. The present sequence is a version of the BASB201 protein
CC sequence of the invention.

XX Sequence 410 AA;

Query Match 7.8%; Score 131; DB 23; Length 410;

Best Local Similarity 24.6%; Pred. No. 0.0078;

Matches 62; Conservative 36; Mismatches 112; Indels 42; Gaps 10;

QY 74 QVKQODPTSKIAQRYGLNWRIGHNNLSSTIYTGQWLTMSGDLKVRERSISSGVT 133

DB 188 QKKQOQALQKQOEHQ-----STINELNKMLADQK-----LNALRANEQALQKQIQR 236

QY 134 AHTSPVAVQSSRPV--QOHPAVQKPTPPVVKKPTPTPPVQOPAPVAPVTEAPPAT 192

DB 237 AEOQAREBEKREKRELAQRQKAEKRT-----SKRYQPTVQGRQLN 278

QY 193 GSSGV---MQRVYVAGTNPVRRFGTATVAGSTVTSNGMFGSDGDLINASNAQVI 248

DB 279 STSGIGAARKQYSLPVSQS--ILHTFG--SIOAGEVRWKGMVIGASAGTPVVAIAAGRVI 334

QY 249 QADH--NMDSASIVIOHTNGFVSSYIHIDQVKTGDTVRTGQRIASMN--QPSGALFE 305

DB 335 LAGYINGGYMIVHGETDLSLYGFQNAVSVKQVLASQVLAQVNTGTEISRSALY- 393

QY 306 FRISNGVYVDP 317

DB 394 FGISRKGTTPNP 405

RESULT 14

AAR85285

ID AAR85285 standard; Protein; 437 AA.

XX AAR85285;

XX 01-APR-1996 (first entry)

DT

XX Lysin.

DE

XX Lysin; autolysin; culture; lactic acid bacteria; fermentation;

KM cheese; foodstuffs; induction.

XX Lactococcus lactis (Strain MG1363).

OS

XX MO9531561-A1.

XX 23-NOV-1995.

PD

XX 12-MAY-1995; 95WO-NL00170.

XX 12-MAY-1994; 94EP-0201353.

XX (UNIL) QUEST INT BV.

XX Buist G, Kok J, Ledebor AM, Venema G;

XX WPI; 1996-010946/01.

DR N-PSDB; AAT06135.

XX Lysis of a culture of lactic acid bacteria in, e.g. cheese
PT production - by in situ prodn. of an auto-lysin, regulated by an
PT inducible promoter.

PS Claim 15; Page 55-57; 103pp; English.

XX In situ production of a homologous autolysin or a heterologous
CC autolysin from a food grade Gram positive bacteria, can be used in
CC a process for the lysis of a culture of lactic acid bacteria. The
CC process can be used in the manufacture of products containing
CC cultures of lactic acid bacteria e.g. cheese, where the culture is
CC lysed following the completion of fermentation. The enhanced
CC induction of the autolysin is performed some hours after the
CC fermentation is finished. No extra lysin needs to be added and the
CC lysin does not need to be isolated or encapsulated. The time of
CC lysis can be precisely controlled.

XX Sequence 437 AA;

Query Match 7.8%; Score 130.5; DB 17; Length 437;

Best Local Similarity 24.1%; Pred. No. 0.0093;

Matches 69; Conservative 39; Mismatches 119; Indels 59; Gaps 12;

QY 26 TTCLAGASAKPTNST-----SGSGSHRTSGSGGLAISQVITDSQVFN- 71

DB 184 TAAITGRVATDPSYGASINRRIISQYNLTRFDGASSAGNTVSG--STTITNNSGTNS 240

QY 72 ---RYOVKQDPTSKIAQRYGLNWRIGHNNLSSTIYTGQWLTMSGDLKVRERSIS 128

DB 241 SSTIYTKSGDITLWGISQRIYSVAQIOSANNLST--IYIGQVLTLGS---ASTNS 295

QY 129 SGVNTAHTSPVAVQSSRPVQOHPAVQKPTPPVVKKPTPT--PPVQOPAPVAPVPT 186

DB 296 GGSNNS-----ASTPTTSVTPA--KPTSGTTVVKVSGDITLWALSVKTKTSLAQLS 345

QY 187 EAPATGSSGVMQ--FRYPVGAATNPVRRFGTATVAGSTVTSNGMFGSDGDLINASN 243

DB 346 WNHLSSDTIYIGQWLIYSQSAASN--STGSGSTATNNNSSTNSNNAISHKV 398

QY 244 AGTVIQADHNMDSASIVIOHTNGFVSSYIHIDQVKTGDTVRTGQ 289

DB 399 KGDITLWGLSQKSGSP-----ASIKAMNH-----SSDTIILIGQ 432

RESULT 15

ABB06928

ID ABB06928 standard; Protein; 518 AA.

XX ABB06928;

XX 18-JUN-2002 (first entry)

DT

XX Micromonospora carbonacea evernimicin locus protein ORF 47.

DE

XX Micromonospora carbonacea; antibiotic; evernimicin; biosynthesis;

KM gene cluster; genetic manipulation; config.

XX Micromonospora carbonacea.

OS

XX WO200155180-A2.

XX 02-AUG-2001.

PD

XX

PF 29-JAN-2001; 2001WO-CA00128.
XX
XX 27-JAN-2000; 2000US-0177711.
XX
XX (ECOP-) ECOPIA BIOSCIENCES INC.
XX
XX (FARN/) FARNET C.
XX
XX Staffa A, Zazopoulos E, Mercure S, Nowacki P;
XX WPI; 2001-476185/51.
XX N-PSDB; ABL50562.
XX
XX Novel isolated gene cluster encoding polypeptides involved in
XX evernomicin biosynthesis useful for construction of evernomicin
XX overproducing strains, and to allow chemical modifications of
XX evernomicin to enhance certain properties -
PS Claim 15; Fig 1; 181pp; English.

CC ABL50555 to ABL50562 represent contigs 1 to 8 from the *Micromonospora*
CC carbonacea evernomicin biosynthetic locus gene cluster. The contigs
CC encode the protein sequences designated ORF (open reading frame) 1 to
CC 49, given in ABL50561 to ABL50590. The gene cluster is useful for the
CC and to allow chemical modifications of evernomicin to enhance certain
CC properties via genetic manipulations of evernomicin to enhance certain
CC gene cluster can be used to produce genetic systems and genes encoding
CC novel enzyme activities and avoid the problems of low yield and quality
CC of evernomicin produced by chemical synthesis.
XX
XX Sequence 518 AA;
SQ

Query Match 7.8%; Score 130.5; DB 22; Length 518;
Best Local Similarity 27.0%; Pred. No. 0.012; Mismatches 112; Indels 53; Gaps 14;
Matches 69; Conservative 22; Mismatches 112; Indels 53; Gaps 14;
DB 20 LIFGTTTCILAGASPTNSTSGS-----SHRTSG-SGGLAI---GQVIT--- 63
QY 8 LAVGTATLVAAGLAGITPFAAAGTAVYVQSGWTFSGNVAITNLSALGWTL 67
DB 64 -----TDSQGVNPRYGV---KQGVYSKIAQRYGLNREIGHINNINSSYTYITQWLT 115
QY 68 TDFPTSGQOVTOGMEATWSQGISVSAAS-----LSWN--GSLGTGSGTIGFNS--- 117
DB 116 WSGDLKVRERSISSGVN-TAHTPSPVAVOSSRPVVOOHDAVOKPTPPVWVKKPTPPV 174
QY 118 WSGSNVPEKSFALNCTGTSVTPTEPTTP-----PTTPTPTPTPTPTPTPTPT 168
DB 175 VQOPAPVAPVTEAPFATGSSGVWQFRYPVGTATNPVRRFGTAIVAG--STVTSNGWFS 232
QY 169 T--PPTTPPTPGAAPLKVSG-----NRLVTASGATYRLIGVNPASGEFACVQGGKMWDS 222
DB 233 GR-DGDLINSMNAQTV 247
QY 223 GPVQASVNMKAMNI 238
DB

RESULT 16
ID ABB54367 standard; Protein; 894 AA.
XX ABB54367;
XX
XX 16-MAY-2002 (first entry)
XX
XX Lactococcus lactis protein p1244.
XX
XX Lactococcus lactis protein p1244.
XX
XX Biosynthesis, biodegradation, lactic bacterium; yogurt; cheese.
XX
XX Lactococcus lactis IL1403.
XX
XX FR2807446-A1.
XX

PD 12-OCT-2001.
XX
XX 11-APR-2000; 2000FR-0004630.
XX
XX 11-APR-2000; 2000FR-0004630.
XX
XX (INRG) INRA INST NAT RECH AGRONOMICQUE.
XX
XX Bolotine A, Sorokine A, Renault P, Ehrlich SD;
XX WPI; 2002-043418/06.
XX
XX New nucleotide sequence useful in the identification of *Lactococcus*
XX lactis and related species -
PS Claim 6; SEQ ID No 1069; 2504pp; French.

CC The present invention is related to a *Lactococcus lactis* nucleotide
CC sequence (AB90521) and related proteins (AB93300-AB95521); the
CC nucleic acid sequence is useful in the detection and/or amplification of
CC related species. The proteins, particularly the invention of *Lactococcus lactis* or
CC biosynthesis or biodegradation of the invention are useful for the
CC production of yogurt and cheese.
CC Note: The sequence data for this patent is based on equivalent patent
CC WO200177334 (published 18-OCT-2001) which is available in electronic
CC format directly from WIPO at http://wipo.int/pub/published_pct_sequences.
XX
XX Sequence 894 AA;
SQ

Query Match 7.8%; Score 130.5; DB 23; Length 894;
Best Local Similarity 23.6%; Pred. No. 0.023; Mismatches 112; Indels 79; Gaps 14;
Matches 69; Conservative 32; Mismatches 112; Indels 79; Gaps 14;
DB 29 ILACGAKPTVNTSGSGSHRTSGSGGLAIGSOVITPDSQVPR--YQVK-----Q 77
QY 620 LLSNAGSEEDANPTDEG-----GRRPFGYGVQWMTDSSGASGRVYMINMTAGVTDN 675
DB 78 GDTVSKIAQRYGLNREIGHINNINSSYTYITQWLTLSGDLKVRERSISSGVYTHPT 137
QY 676 PDIIT--AQFKLWM-----HNQ-----GOWTAKSYYSWTQPMULININIA--- 717
DB 138 SPVAVOSSRPVVOOHPA-----YOKPTPPVYVVKETPTPPVQOPAPVAPVTEAPATG 193
QY 718 TQAFVANFERDLNGHPEBSTWAGBNKFNKIPSGGGGYI---APISPII---VTS 770
DB 194 SSGVWQFRYPVGTATNPVRRFGTAIVAGSTVTSNGMFWSGRGLIN-----ASNAG 245
QY 771 EMG-----WTSPTTGAQGFHAM-----DLVNGNPTTPILASGIG 806
DB 246 TVIQADHN--MDGASIVIOHTNGFVSYYTHIKDAQYKTDVYRTGORTASM 234
QY 807 QVQVAGSNYYDWYGVYTVIKHADGLYGVYAGSRIDVSVGVQVKKGGQIGLM 858
DB

RESULT 17
ID AAO17657 standard; Protein; 410 AA.
XX AAO17657;
XX
XX 05-AUG-2002 (first entry)
XX
XX H influenzae BASB201 #2.
XX
XX BASB201; otitis media; pneumonia; sinusitis; nosocomial infection;
XX
XX auditive nerve damage; delayed speech learning; vaccine;
XX
XX antibacterial; auditory; antiinflammatory.
XX
XX Haemophilus influenzae.
XX
XX
XX Key Location/Qualifiers

```

FT Domain 32..300
FT /label= KEQ_rich_domain
FT Domain 301..400
FT /label= peptidase_M37-like_domain
XX
XX WO200230967-A2.
XX
XX 18-APR-2002.
XX
XX 05-OCT-2001; 2001WO-EP11561.
XX
XX 13-OCT-2000; 2000GB-0025169.
XX
XX (SMK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
XX Thonard J;
XX
XX WPI; 2002-426267/45.
XX
XX N-PSDB; AAL46625.
XX
XX New isolated non-typable Haemophilus influenzae BASB201 polypeptides,
XX useful as components of vaccines for treating bacterial infection such
XX as otitis media, delayed speech learning and inflammation of middle ear
XX
XX Claim 3; Page 88; 90pp; English.
XX
XX The present invention provides the protein and coding sequences of
XX several versions of the BASB201 protein from non-typable Haemophilus
XX influenzae. These can be used in the production of vaccines against H.
XX influenzae infection, which can cause otitis media in infants and
XX children, pneumonia in elders, sinusitis, nosocomial infections, or
XX invasive diseases, chronic otitis media with hearing loss, fluid
XX accumulation in the middle ear, auditive nerve damage, delayed speech
XX learning, infections of the upper respiratory tract and inflammation of
XX the middle ear. The present sequence is a version of the BASB201 protein
XX sequence of the invention.
XX
XX Sequence 410 AA;
XX
XX Query Match 7.8%; Score 130; DB 23; Length 410;
XX Best Local Similarity 25.2%; Pred. No. 0.0094;
XX Matches 64; Conservative 33; Mismatches 111; Indels 46; Gaps 11;
XX
XX 74 QVKQSDTVSKIAQRGLMREIGHNNINSSITTYTGQMLTMSDGLKVRRESISSGVNT 133
XX ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 188 QKKQQQALQKQKQEHQ-----STLNELNKNLALDQDK-----INALKANEQALKQEIQR 236
XX
XX 134 AHTSPVAVQSSRP---VQCHPAVQKPTPPVAVVKKPTPPPVVQOPAPVAPVTEAPF 190
XX ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 237 AE--QAVNEQEKREHQAQKQKAEKRT-----SKPYQPTVQERQL 276
XX
XX 191 ATGSSGV---MOPRYPVGATNPVVRFGTATVAGSTVTSNGMWSGDDGLINASNAGT 246
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 277 INSTGLGAQKQYSLPVSGS--ILHTFG--SIQGEVRMKGMVIGASAGTPVKAIAGR 332
XX
XX 247 VTQANH-NMDGASTIYIQTNGFVSYYIHAKAQVTVGTGQIJAAMKN--QSSGAL 303
XX ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 333 VILAGYLVNGYGVMTVKHGETDLSLYGFNQAVSVKVGDLVSAQVIAOVGMTGEISRSL 392
XX
XX 304 FEFRIKNGVYVDP 317
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
XX 393 Y-FGISRKGTVPNP 405
XX
XX RESULT 18
XX ID AAY31384 standard; Protein; 1708 AA.
XX AC AAY31384;
XX
XX 12-OCT-1999 (first entry)
XX

```

```

DE HEV-US2 ORF1 protein.
XX
XX Hepatitis E virus; HEV; binding partner; virus; US-HEV infection;
XX vaccine; passive immunisation.
XX
XX Hepatitis E virus.
XX
XX Key Location/Qualifiers
XX
XX Misc-difference 322
XX /label= unknown
XX /note= "encoded by CRG"
XX
XX Misc-difference 331
XX /label= unknown
XX /note= "encoded by RAC"
XX
XX Misc-difference 445
XX /label= unknown
XX /note= "encoded by RAC"
XX
XX Misc-difference 448
XX /label= unknown
XX /note= "encoded by RAC"
XX
XX Misc-difference 634
XX /label= unknown
XX /note= "encoded by RTG"
XX
XX Misc-difference 646
XX /label= unknown
XX /note= "encoded by CCY"
XX
XX Misc-difference 811
XX /label= unknown
XX /note= "encoded by GCS"
XX
XX Misc-difference 1533
XX /label= unknown
XX /note= "encoded by TKT"
XX
XX Misc-difference 1578
XX /label= unknown
XX /note= "encoded by TAY"
XX
XX Misc-difference 1691
XX /label= unknown
XX /note= "encoded by RGC"
XX
XX Misc-difference 1691
XX /label= unknown
XX /note= "encoded by GAR"
XX
XX WO9919732-A1.
XX
XX 22-APR-1999.
XX
XX 15-OCT-1998; 98WO-US21941.
XX
XX 15-OCT-1997; 97US-0061199.
XX
XX (ABBO ) ABBOTT LAB.
XX
XX Dawson GJ, Desai SM, Erker JC, Mushahwar IK, Schlauder GG;
XX WPI; 1999-288017/24.
XX
XX N-PSDB; AAZ00267.
XX
XX Detection of United States isolates of hepatitis E virus
XX
XX Claim 6; Page 216-221; 260pp; English.
XX
XX The invention provides a method for detecting a US (sub)type hepatitis E
XX virus (US-HEV), or its naturally occurring variants in a sample by
XX treatment with a binding partner specific for a marker of the virus, and
XX then detecting any complex formed. The method is used to diagnose
XX infection with US-HEV. Polypeptides from US-HEV, antibodies specific for
XX open reading frames (ORF) in US-HEV and host cells expressing these ORFs
XX are useful in vaccines or for passive immunisation. The polypeptides are
XX also used to raise specific antibodies (useful as immunoassay reagents).
XX Fragments of nucleic acid from US-HEV are useful as primers and probes in
XX usual hybridisation and amplification assays for detecting infection. The
XX present sequence represents a HEV-US2 ORF1 protein.
XX
XX Sequence 1708 AA;
XX
XX Query Match 7.6%; Score 127; DB 20; Length 1708;
XX

```

Best Local Similarity 22.7%, Pred. No. 0.11;
Matches 80; Conservative 37; Mismatches 120; Indels 116; Gaps 19;

QY 27 TCILAGC---ASKRTYSTSGSGHRTSGSGGLAI-----GQVITTSQVPRRYQVK 76
DB 574 ITVDGAHLNAGEBEVYLSFDASRQSMGAGSHSLTYELTPAGLOVKISSNGLDCTATFP 633
QY 77 QGDIVS-----KIAQRGLN---WR-----EIGHI---NN 100
DB 634 XGAPSAAPGEVXAFCSALYRYNFTQRHSLTGMLMHPGLGIPPPSPGHIMESANP 693
QY 101 INSSYTYTGOMTLI-NSGDLKVERSSISGVNTAHTPSPFAVQSSRPV----- 149
DB 694 FCGEGTLVTRTWSTSGFSSDPSPEEAPAPASAAAPGLPYPT-----PVSIDIWLPSPS 747
QY 150 -QQH-----PAYQKP---TPPVVYVKKPTPTPPVVOQAPAPVAPVTEAPFATSSGVWQ 199
DB 748 EESHVDASVSPSEBPAGLTSPIVLT--PPPPPPVAKRATSPFPATRR-----IL 796
QY 200 FRYPVGATNPVVRFRGTATVAGSTVTSNGMFMFGSDGDLINASNAGTVIQADHNDGAST 259
DB 797 YTPPDGA-----KYVAGSLXESDCDM-----LVNASNPG-----HRPGCG-- 831
QY 260 VIOHTNGFVSSYIHIKDAQVKTGDTVRTGQRIAMKNQP--SGAALFEPRISHN 311
DB 832 -LCH-----AFYQRFPEAFYSTERIMEGLAAYLTLPRIHADVADYVEQN 878

RESULT 19
AAR24147
ID AAR24147 standard; Protein, 374 AA.
AC AAR24147;
XX
XX 15-NOV-1992 (first entry)
DT
XX
XX Beta-lytic protease.
DE
XX Gram-negative; Gram-positive; bacteria; decomposition.
OS
XX Achromobacter lyticus.
PN JP04108387-A.
XX
XX 09-APR-1992.
PD
XX 29-AUG-1990; 90JP-0225136.
PF
XX 29-AUG-1990; 90JP-0225136.
PR
XX 29-AUG-1990; 90JP-0225136.
PA (WAKP) WAKO PURE CHEM IND.
XX
XX WPI; 1992-171653/21.
DR N-PSDB; AAQ25083.
XX
XX The protein sequence of the beta lytic peptidase from Achromobacter
CC lyticus was deduced from the DNA sequence obtd. by PCR using
CC primers based on the sequence of the L. enzymogenes beta-protease.
CC The beta-lytic protease is expected to be an enzyme which can decompose
CC not only Gram-positive bacteria but also some Gram-negative bacteria.
XX
XX
PS Claim 2; Fig 1; 13pp; Japanese.
PT
PT Beta-lytic protease gene and DNA encoding it - for decomposing
PT Gram-positive and some Gram-negative bacteria
XX
XX
SQ Sequence 374 AA;
Query Match 7.6%; Score 126.5; DB 13; Length 374;
Best Local Similarity 25.9%; Pred. No. 0.017;
Matches 50; Conservative 29; Mismatches 65; Indels 49; Gaps 7;

DB 154 AGRAAARRRRVAGL-----RPYQRTAPGGGGRGP-----LR 186
QY 177 QAPAPAPVTEAPFATSSGVWQFRYVGTNPVVRFRGTATVAGS-----YVTSNG 228
DB 187 QGRP-----GRAVSPNGLQFPFPGASWHV---GGAHNTSGCNTPMSSLDMSRGG 236
QY 229 MFMFGSDGDLINASNAGTVIQADHNDGASIVIOHTNGFVSSYIHIKDAQVKTGDTVRTG 288
DB 237 GWSGNQNNWVASASAGSFKR--HSCFAEIV--HTGWSYTYHLMNIQYNTGANVSMN 292
QY 289 QRIAMKNQPSGA 301
DB 293 TAIANPANTQQA 305

RESULT 20
AAY44646
ID AAY44646 standard; Protein, 374 AA.
AC AAY44646;
XX
XX 18-APR-2000 (first entry)
DT
XX
XX Achromobacter lyticus beta-lytic protease.
DE
XX Beta-lytic protease; antibiotic; anti-staphylococcal;
KM Gene therapy; mastitis; staphylococcal infection; ruminant; cow;
KM transgenic animal; altered beta-lytic protease gene.
XX
XX
OS Achromobacter lyticus.
PN WO967381-A1.
XX
XX 29-DEC-1999.
PD
XX 22-JUN-1999; 99WO-US14073.
PF
XX 22-JUN-1998; 98US-0090175.
PR 21-JUN-1999; 99US-0337079.
XX
XX (UYVE-) UNIV VERMONT & STATE AGRIC COLLEGE.
PA
XX Bramley JA, Plant KI, Kerr D;
PI
XX WPI; 2000-147208/13.
DR N-PSDB; AA249721.
XX
XX Treatment of Staphylococcal infections, such as mastitis, in ruminant
PT animals, especially cows -
PT
XX
XX Disclosure; Fig 14B; 61pp; English.
PS
XX The present sequence is a Achromobacter lyticus beta-lytic protease
CC which has anti-staphylococcal activity. The beta-lytic protease gene is
CC used in the production of altered genes which
CC allow expression and preferably secretion of active protein in mammalian
CC cells/tissues. The altered gene is produced by operably linking the
CC beta-lytic protease coding sequence with mammalian promoter, signal
CC peptide and translation initiation sequences. The modified sequence
CC is used in gene therapy to treat staphylococcal mastitis infections in
CC ruminants, e.g. goats, sheep, and cows. It is also used to produce
CC transgenic animals which are resistant to staphylococcal infections.
XX
XX
SQ Sequence 374 AA;
Query Match 7.6%; Score 126.5; DB 21; Length 374;
Best Local Similarity 25.9%; Pred. No. 0.017;
Matches 50; Conservative 29; Mismatches 65; Indels 49; Gaps 7;

QY 177 QPAPVAPVTEAPFATGSSGVMQFRYPVG--ATNPVRRFGTATVAGS-----TGTENG 228
DB 187 QGRP-----GRAVSPNGILQFPFRGASWHV---GGATHTGSSGNYPMSLDMSCG 236
QY 229 MWFSGRGGDLINASNAGTVIQADHNMDSASIVIOHTNGFVSSYHIKDAQVKTGDTV 288
DB 237 GWSGNQNNWVWSASASAGSFKR--HSSCFAEIV--HTGWSSTYYHLMNIQVNTGANVSMN 292
QY 289 QRASMKQPSGA 301
DB 293 TAIANPANTQAQA 305

RESULT 21
AAG92487
ID AAG92487 standard; Protein; 205 AA.
AC AAG92487;
XX
XX
DT 26-SEP-2001 (first entry)
XX
DE C glutamicum protein fragment SEQ ID NO: 6241.
XX
XX Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KW organic acid synthesis.
XX
XX Corynebacterium glutamicum.
XX
XX EP1108790-A2.
XX
XX 20-JUN-2001.
XX
XX 18-DEC-2000; 2000EP-0127688.
XX
XX 16-DEC-1999; 99JP-0377484.
XX 07-APR-2000; 2000JP-0159162.
XX 03-AUG-2000; 2000JP-0280988.
XX
XX (KYOW) KYOWA HAKKO KOGYO KK.
XX
XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
XX WPI: 2001-376931/40.
XX N-PSDB; AAH67706.
XX
XX Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
PT
XX
XX Claim 17; SEQ ID NO: 6241; 246bp + Sequence Listing; English.
XX
XX The present invention provides a number of nucleotide and protein
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of coryneform bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Coryneform bacterium, and identifying a homologue of a gene derived
CC from coryneform bacterium. Coryneform bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a protein described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
XX Sequence 205 AA;
SQ

Query Match 7.5%; Score 125.5; DB 22; Length 205;
Best Local Similarity 33.8%; Pred. No. 0.0093;
Matches 52; Conservative 14; Mismatches 73; Indels 15; Gaps 5;

QY 179 APVAPVTEAPFATGSSGVMQFRYPVG--ATNPVRRFGTATVAGS-----TGTENG 226
DB 41 ATMAPASAGTITVAGSSGVADTVAAGVATTAAPATVAPRANGFTSGFGRMGTFH 100
QY 227 NGMWFSGRGGDLINASNAGTVIQADHNMDSASIVIOHTNGFVSSYHIKDAQVKTGDTV 285
DB 101 NGDIANSIGPIYAVWAGTVISSGPASGVGMQRIHDDGSISYIGHMEYLYVSGERV 160
QY 286 RTGQRASMKQPSGAALFEFRISRGVY-VDP 317
DB 161 AAGQELAGMSQGFSTGSHLFEIHPDGVTPVDP 194

RESULT 22
AAW95501
ID AAW95501 standard; Protein; 288 AA.
AC AAW95501;
XX
XX
DT 26-MAR-1999 (first entry)
XX
DE B. subtilis Yoch related gene product YochBsubt.
XX
XX RP-factor; resuscitate; latent cell; growth-stimulation; receptor; YABE;
KW convertase; vaccine; immunotherapy; prophylaxis; excipient; antibiotic;
KW antimicrobial; tuberculosis; food; pharmaceutical; culture additive.
XX
XX Bacillus subtilis.
XX
XX WO9855624-A1.
XX
XX 10-DEC-1998.
XX
XX 03-JUN-1998; 98WO-GB01619.
XX
XX 27-MAY-1998; 98GB-0011221.
XX 04-JUN-1997; 97GB-0011389.
XX
XX (UYWA-) UNIV WALES.
XX
XX Kaprelyants AS, Kell DB, Mukamolova GV, Young DI;
PI Young M;
XX
XX WPI: 1999-070270/06.
XX
XX New bacterial resuscitation factors - useful for stimulating latent
PT bacteria and growth, particularly for use as antimicrobials, also
PT vaccines comprising bacteria with mutations in resuscitation factor
PT genes
XX
XX Disclosure; Fig 1B; 76pp; English.
XX
XX The invention relates to RP-factors which are substances that can
CC resuscitate dormant, moribund or latent cells, possibly also having
CC growth-stimulating activity. Host cells containing a vector comprising
CC the RP-factor or its receptor encoding nucleic acid can be used for the
CC recombinant production of the RP-factor. RP-factors, their receptors or
CC convertases, antibodies (against the RP-factors and RP-factor receptors
CC or convertases), antagonists and agonists, are useful in vaccines and for
CC immunotherapy, diagnosis and prophylaxis. They are also useful as
CC excipients, generally as antimicrobials and especially for infections
CC associated with latency. They can be used as potentiators of antibiotics
CC such as isoniazid, streptomycin etc., in treatment of tuberculosis.
CC RP-factor can also be used: to determine the microbiological quality of
CC foods, pharmaceuticals, medical products; as culture additive for
CC bacteria; to stimulate growth and/or to resuscitate microorganisms and to
CC produce libraries of biomolecules and microorganisms (which may then be
CC screened for useful products). Probes complementary to the RP-factor
CC nucleic acid are used identify and clone other RP-factor genes. The
CC inventions may also be used to prevent bacterial resuscitation. Breaking
CC dormancy with RP-factor facilitates detection, culture and enumeration of
CC many bacteria. The present sequence represents a B. subtilis Yoch related

CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
 CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.

XX Sequence 639 AA;

Query Match 7.2%; Score 120.5; DB 22; Length 639;
 Best Local Similarity 21.7%; Pred. No. 0.11;

Matches 54; Conservative 32; Mismatches 84; Indels 79; Gaps 7;

QY 24 VITTCILAGCASKPTYNSTSGSGSHRTSGGGLAIGSVITTSOG----- 68
 DB 432 VITKRLILSSASSPFIQSSYGVGINQRFHAAHKKQSQDDQDQASGLQSPPSRLSPTL 491
 QY 69 -----VNNRYGVKQG--DVSKIAQRYGLMREIGHININLSYTIYTG 110
 DB 492 IDNSAKQLANTVTVQLSRFTSQGPIKPVSPNSPFGTDRNLNANR----- 543
 QY 111 QMLTMSGDLKVRKRSISGVNTATPPSPVAVQSRPPVQCHPAVQKTPPVVVVKKPTP 170
 DB 544 -----GD-----TSHSPTPGKVSPLSPLS--PGIKSPITIPAEKGNPP 581
 QY 171 TPVPVQAPVAPVTEAPFATGSSGVMQFRYPVGTATNPVVRFGTATVAGSTVTSNGMW 230
 DB 582 IP-----PKKPGILTPSPSAT-----TPLTKTHSQAASTLTAAEDLASSCSSTIV 625
 QY 231 FSGRDGLI 239
 DB 626 ANGVVELL 634

RESULT 25

AAE18364
 ID AAE18364 standard; Protein; 357 AA.

XX AC AAE18364;

XX AC AAE18364;

DT 07-MAY-2002 (first entry)

XX Streptococcus pyogenes strain B514 BVH-PI mature protein.

XX BVH-PI gene; streptococcal infection; pharyngitis; erysipelas; impetigo;
 KM scarlet fever; bacteraemia; necrotising fasciitis; toxic shock; vaccine;
 KM immune response; anti-inflammatory; immunisation; antibacterial.

XX OS Streptococcus pyogenes B514.

XX PN WO200204495-A2.

XX PD 17-JAN-2002.

XX PF 06-JUL-2001; 2001WO-CA01001.

XX PR 06-JUL-2000; 2000US-216465P.

XX PA (SHIR-) SHIRE BIOCHEM INC.

XX PI Martin D, Hamel J, Brodeur B;

XX DR WPI; 2002-171701/22.

XX N-PSDB; AAD29299.

XX New Streptococcus pyogenes antigen useful for diagnosing, preventing or
 PT treating streptococcal infection, e.g. pharyngitis, erysipelas and
 PT impetigo, scarlet fever, and invasive diseases

XX Claim 21; Fig 16; 74pp; English.

XX The invention relates to antigens, more particularly an antigen of

CC Streptococcus pyogenes (also called group A Streptococcus (GAS))
 CC bacterial pathogen. The polypeptides and polynucleotides encoding them
 CC are useful for diagnosing, preventing or treating streptococcal
 CC infection, such as pharyngitis, erysipelas, impetigo, scarlet fever,
 CC invasive diseases (bacteraemia, necrotising fasciitis, toxic shock), and
 CC for eliciting an immune response. The polypeptides may also be used as
 CC immunogens for producing antibodies for the diagnosis and treatment of
 CC Streptococcus infection, or for passive immunisation. DNAs encoding
 CC polypeptides may also be used to design DNA probes for detecting the
 CC presence of Streptococcus in biological samples suspected of containing
 CC the bacteria. The vaccine composition is useful as a prophylactic or
 CC therapeutic treatment of Streptococcal infection in an individual
 CC susceptible to or infected with streptococcal infection. The present
 CC sequence is Streptococcus pyogenes strain B514 BVH-PI mature protein.

XX Sequence 357 AA;

Query Match 7.2%; Score 120; DB 23; Length 357;
 Best Local Similarity 25.8%; Pred. No. 0.056;

Matches 61; Conservative 21; Mismatches 78; Indels 76; Gaps 11;

QY 73 YQVQSGDPTVSKIAQRYGLNMR---EIGHININLSY--TIYT-----GQMLTMSGDLK 121
 DB 24 YTVXYGDTLSTIAAMGIDVAVLDINHIANIDIFPDTITANYNOHGQATTL----- 77
 QY 122 VRESISGVNTATHTPS---PVAVOSRPPVQCHP-----AVOKPFPVVVVKKKPTP 171
 DB 78 TVQAPASSPASVSHVSPSEPIPOASATSQPVPMAFKAFTPLASAKPDSSTVASELST 137
 QY 172 -----PPVQO--PAPVAPVTEAPFATGSSGVMQFRYPVGTATNPVVRFG 215
 DB 138 NDVSTESSESQKQPEVQEAVPFPKAAETTEVEPKTDIS-----EDPTSANRPVPNESA 192
 QY 216 TATVAG-----STVTSNGMFFSGRDGLIASNAG 245
 DB 193 SEEVSAAPAPAKPKETSAAPAQAVADTTVATSNGLSYA--PNHAYPNMNA 246

RESULT 26

AAE18360
 ID AAE18360 standard; Protein; 382 AA.

XX AC AAE18360;

DT 07-MAY-2002 (first entry)

XX Streptococcus pyogenes strain B514 BVH-PI protein.

XX BVH-PI gene; streptococcal infection; pharyngitis; erysipelas; impetigo;
 KM scarlet fever; bacteraemia; necrotising fasciitis; toxic shock; vaccine;
 KM immune response; anti-inflammatory; immunisation; antibacterial.

XX OS Streptococcus pyogenes B514.

XX PN WO200204495-A2.

XX PD 17-JAN-2002.

XX FT Key Location/Qualifiers
 FT Peptide 1..25
 FT Protein /label= Signal_peptide
 FT /note= "Mature_BVH_PI_protein"

XX OS Streptococcus pyogenes B514.

XX PN WO200204495-A2.

XX PD 17-JAN-2002.

XX PF 06-JUL-2001; 2001WO-CA01001.

XX PR 06-JUL-2000; 2000US-216465P.

XX PA (SHIR-) SHIRE BIOCHEM INC.

XX PI Martin D, Hamel J, Brodeur B;

XX WPI; 2002-171701/22.

XX Human, secreted protein; cancer; tumour; cardiovascular disorder;
KM blood disorder; haemophilia; autoimmune disease; diabetes; inflammation;
KM infection; fungal; bacterial; viral; HIV; allergy; arthritis;
KM neurodegenerative disease; asthma; contraceptive; open reading frame;
ORF.
XX Homo sapiens.
OS
PN WO200011015-A1.
PD 02-MAR-2000.
XX
PF 24-AUG-1999; 99WC-US19351.
XX
PR 24-AUG-1998; 98US-0097638.
PR 24-AUG-1998; 98US-0097659.
PR 03-SEP-1998; 98US-0098618.
PR 28-SEP-1998; 98US-0102092.
PR 25-NOV-1998; 98US-0109978.
PR 23-DEC-1998; 98US-0113645.
PR 23-DEC-1998; 98US-0113646.
PR 23-AUG-1999; 99US-0379246.
XX
PA (ALPH-) ALPHAGENE INC.
XX
PI Valenzuela D, Yuan O, Hoffman H, Hall J, Raplejo P;
XX
DR WPI; 2000-224657/19.
XX
PI New secreted or transmembrane proteins and polynucleotides encoding
PT them, useful for treating neurodegenerative disorders, autoimmune
PT diseases and cancer -
XX
PS Disclosure; Page 350-351; 357pp; English.
XX
XX The invention relates to 40 human secreted proteins (AA94981-Y95020),
CC and CDNA sequences encoding them (AA23423-A23462). The secreted
CC proteins of the invention include those that are thought to be only
CC partially secreted, i.e., transmembrane proteins. The proteins of the
CC invention may exhibit one or more activities selected from the following:
CC cytokine activity; cell proliferation; differentiation; immune
CC modulation; haematopoiesis regulation; tissue growth activity;
CC actin/inhibin activity; chemotactic/chemokinetic activity; haemostatic
CC and fibrinolytic activity; anti-inflammatory activity; and tumour
CC inhibition activity. The proteins may be administered to patients as
CC vaccines, and the nucleotides may be used as part of a gene therapy
CC regime. Diseases or conditions that may be treated using the proteins or
CC nucleotides of the invention include autoimmune diseases; genetic
CC disorders; haemophilia; cardiovascular diseases; cancer; bacterial,
CC fungal and viral infections, especially HIV; multiple sclerosis;
CC rheumatoid arthritis; pulmonary inflammation; Guillain-Barre syndrome;
CC insulin dependent diabetes mellitus; and allergic reactions such as
CC asthma and anaemia. They may also be used for treating wounds, burns,
CC ulcers, osteoporosis, osteoarthritis, periodontal diseases, Alzheimer's
CC disease, Parkinson's disease, Huntington's disease and amyotrophic
CC lateral sclerosis (ALS). Proteins with actin/inhibin activity may
CC additionally be useful as contraceptives. Nucleic acid sequences of the
CC invention may be used in chromosome mapping, and as a source of the
CC diagnostic primers and probes. Sequences AA95021-Y95023, AA95026-Y95028
CC and AA95030 represent additional open reading frames (ORFs) encoded by
CC the CDNA clones of the invention.
XX
SQ Sequence 284 AA;
Query Match 7.1%; Score 118; DB 21; Length 284;
Best Local Similarity 39.6%; Pred. No. 0.062;
Matches 36; Conservative 8; Mismatches 33; Indels 14; Gaps 6;

OY 132 NTAHAPSPV---AVSSRPVQGHAYOKPPTVVVK---KPTPPVQVQAPAPAPV 185
DB 115 STVPABSPLSAAVAVBSKLPEDDEPPAPPPPPASVSPQAEVPTTPA---PAPAPPS 171

OY 186 T-EAPATGSSGWMQ---FRYPVGTNPVVR 212
DB 172 TPAAPKRRGSSGVDETLFALP-AASEVVR 201

RESULT 31
ID AAM52321 standard; Protein; 633 AA.
AC AAM52321;
XX 18-JAN-2002 (first entry)
DE Las17 protein.
XX
KM Actin polymerisation; Ena/VASP, vasodilator-stimulated phosphoprotein;
KM metastatic cancer; parasitic infection; cytotoxic; Las17.
OS Saccharomyces cerevisiae.
XX
PN WO200171356-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001, 2001WO-FR00843.
XX
PR 22-MAR-2000; 2000FR-0003637.
XX
PA (CNRS) CENT NAT RECH SCI.
PA (CURRI-) INST CURIE.
PI Fradelizi J, Friederich E, Golsteyn RM, Louvard D, Noireaux V,
PI Sykes C;
XX
DR WPI; 2001-639148/73.
XX
PT Identifying modulators of actin polymerization, potentially useful for
PT treating tumor metastasis and parasitic infection, using proteins that
PT contain Ena/VASP binding sites -
XX
PS Claim 13; Pages 105-107; 109pp; French.
XX
XX The present invention relates to a method for identifying modulators of
CC actin polymerization. The method involves using proteins that contain at
CC least one binding motif for proteins of the Ena/VASP
CC (vasodilator-stimulated phosphoprotein) family in the preparation of
CC reagents for identification/screening of molecules that modulate
CC formation of the actin cytoskeleton. The proteins used in the method
CC (i.e. the proteins with binding motif(s) for Ena/VASP proteins) do not
CC bind to the Arp2/3 protein complex. The modulators identified by the
CC method are potentially useful for treating disorders of actin
CC polymerisation, e.g. metastatic cancer or parasitic infection, and as
CC cytotoxic agents. The present sequence one such protein with binding
CC motif(s) for Ena/VASP proteins, which was used in the method of the
CC present invention.
XX
SQ Sequence 633 AA;
Query Match 7.1%; Score 118; DB 22; Length 633;
Best Local Similarity 23.0%; Pred. No. 0.17; Indels 80; Gaps 12;
Matches 65; Conservative 34; Mismatches 104;

OY 37 PYNSTSGSG-----SHRTSGGAIAGSVITDSOG----- 68
DB 329 PMRTTSGSGVRLPAPPPPPRGPAPPPPHRVHTSNTLNSAGNSLLPQATGRGPPAP 388

OY 69 -VNRVQVQGGTGVKINQRYGLMWREIGHNNINS-----YTIYQGMWLTMSG 118
DB 389 PPRPARPTPNTVMQNNQVQNNNSNRPGVQTNNSMSPPPVPTTNTLTTPQ-WTAAG 447

OY 119 DLKY-----KRSISGVNTAHTPSPVAVSSRPVQGHAYOKPPTVVVKETPTTP 172
DB 448 QPAPVLPONTQAPSOATNVVAPPPPPASLSQSOIPQASAPAPLPILBETTSIADPPPP 507

QY 173 P-VVOQP-----APVAPPTEAPFATGSSGVMQFRYPVGNTPVRRFGTAVAGSTVSN 227
 DB 508 AFLTQPOGSGAPAPPPPPQMP-ATSTSG-----GGSFAETT 543
 QY 228 GWFSGRDGDLIN---ASNAGTVIQADHN-MDGASIVIQHTNG 266
 DB 544 G--DAGRALLASIRGAGGIGALRKVDKSQLDKPSVLLQEARG 584

RESULT 32
 AAG67365
 ID AAG67365 standard; Protein; 633 AA.
 AC AAG67365;
 DT 13-NOV-2001 (first entry)
 DE Amino acid sequence of a yeast Las17 protein.
 KW Wiskott-Aldrich syndrome protein; WASP; actin cytoskeleton; Las17;
 KM cell motility; actin polymerisation; cancer; parasite infection;
 OS embryonic development; immune response; wound repair.
 XX Saccharomycetes cerevisiae.
 PN MO200144292-A2.
 XX 21-JUN-2001.
 PD 15-DEC-2000; 2000WO-FR03569.
 PE 16-DEC-1999; 99FR-0015900.
 PR (CNRS) CENT NAT RECH SCI.
 PA (CURIE-) INST CURIE.
 XX Noireaux V, Prost J, Sykes C, Friederich E, Golsteyn RM;
 PI Louvard D;
 DR WPI; 2001-536241/59.
 XX N-PSDB; AAH77917.
 PT New fragments of WASP family proteins, useful for detecting and
 PT identifying modulators of actin cytoskeleton formation, potential
 PT anticancer and antiparasitic agents -
 XX Claim 13; Fig 7; 162bp; French.
 PS The present sequence represents a Las17 protein. Las17 is a member of
 CC the WASP (Wiskott-Aldrich syndrome protein) family of proteins.
 CC Peptide fragments of WASP-family proteins of eukaryotic cells are used
 CC to prepare reagents for detecting compounds that inhibit or stimulate
 CC formation of the actin cytoskeleton, and thus inhibit or stimulate cell
 CC motility. The peptides are used to detect and identify compounds which
 CC are potentially useful for treating diseases associated with dysfunction
 CC of actin polymerisation, particularly metastatic cancer and parasite
 CC infection; as cytotoxic agents for inhibiting/stimulating formation of
 CC the actin cytoskeleton and for detecting side-effects, on actin
 CC polymerisation, of pharmaceuticals. By modulating actin polymerisation,
 CC these compounds affect cell motility, embryonic development, the immune
 CC response and wound repair.
 XX Sequence 633 AA;
 SQ

Query Match 7.1%; Score 118; DB 22; Length 633;
 Best Local Similarity 23.0%; Pred. No. 0.17;
 Matches 65; Conservative 34; Mismatches 104; Indels 80; Gaps 12;

QY 37 PLYNTSSG-----SHRTSGSGALATGQVTTDSG-----68
 DB 329 PWRITSGSVRLPAPPPRRGAPAPPPPHRHVTSNTLNSAGNSLLPQATGRGAPAP 388

QY 69 -VPRRYQVKQDPTVSKIAQRYGLNWRREIGHINNLS-----YTIYTGQMLTLMSG 118
 DB 369 PPRASRRTPTVWTMQNQNNNSRPFQYQTNMSMSBPPEPVTTFTMLTPQ-MTAAFG 447
 QY 119 DLKV-----RERSISGVNTAHTPSPVAVOSSRPVQHPAVQKPTPPVVVKKPTPT 172
 DB 448 QPAVPLPONTQAPSGATVVPVAPPPPPASLQSQIPQSPAPAPPTLPSTTSAPPPPP 507
 QY 173 P-VVOQP-----APVAPPTEAPFATGSSGVMQFRYPVGNTPVRRFGTAVAGSTVSN 227
 DB 508 AFLTQPOGSGAPAPPPPPQMP-ATSTSG-----GGSFAETT 543
 QY 228 GWFSGRDGDLIN---ASNAGTVIQADHN-MDGASIVIQHTNG 266
 DB 544 G--DAGRALLASIRGAGGIGALRKVDKSQLDKPSVLLQEARG 584

RESULT 33
 AAY56967
 ID AAY56967 standard; Protein; 1192 AA.
 AC AAY56967;
 DT 25-APR-2000 (first entry)
 DE Human MAGI polypeptide.
 KW MAGI protein; neuroendocrine-specific protein; neuropathy; human;
 KM spinal injury; neuronal degeneration; neuromuscular disorder; cancer;
 KM psychiatric disorder; developmental disorder; inflammatory disorder;
 KM stroke; cytostatic; cerebroprotective; neuroprotective.
 OS Homo sapiens.
 XX WO200005364-A1.
 EN 03-FEB-2000.
 PD 21-JUL-1999; 99WO-GB02360.
 PE 22-JUL-1998; 98GB-0016024.
 PR 19-JUL-1999; 99GB-0016898.
 XX (SMIK) SMITHKLINE BEECHAM PLC.
 PA Michalovich D, Prinjha RK;
 PI WPI; 2000-182693/16.
 DR N-PSDB; AAZ56886.
 XX Novel polypeptides related to neuroendocrine-specific proteins and
 PT polynucleotides useful for diagnosis of various diseases and for
 PT treatment of cancer and neurological disorders -
 XX Claim 2; Page 20-21; 35pp; English.
 PS The invention relates to human MAGI protein, which is similar to
 CC neuroendocrine-specific protein. The MAGI protein can be expressed by
 CC standard recombinant methodology. The MAGI polypeptides, polynucleotides
 CC and antibodies are useful for treating diseases, including neuropathies,
 CC spinal injury, neuronal degeneration, neuromuscular disorders,
 CC psychiatric disorders and developmental disorders, cancer, stroke and
 CC inflammatory disorders. The polynucleotide is also useful for chromosome
 CC localization and for tissue expression studies. The present sequence
 CC represents the human MAGI protein.
 XX Sequence 1192 AA;
 SQ

Query Match 7.1%; Score 118; DB 21; Length 1192;
 Best Local Similarity 39.6%; Pred. No. 0.33;
 Matches 36; Conservative 8; Mismatches 33; Indels 14; Gaps 6;

QY 132 NTAAHPSFV---AVOSSRPVQHPAVQKPTPPVVVVK---KPTPTPVVQAPVAPV 185

DB	Accession	Protein Name	Location/Qualifiers
DB	115	STVPAAPSLTAAVSPSKLPEDDEPPARPPPPASVSPQAEVWTPPA---PAPAAPFS 171	
QY	186	T-EAEPFAGSSGVQ---FRYPGATNPVR 212	
Db	172	TPAAPKRGSSGSVDETLFLALF-AASEPVR 201	
RESULT 34			
AAU04591			
ID	AAU04591	standard; Protein; 1192 AA.	
XX			
AC	AAU04591;		
XX			
DT	26-SEP-2001	(first entry)	
XX			
DE		Human Nogo protein.	
XX			
KW		Human; Nogo receptor; axonal growth; immunogen; antibody; nogo protein;	
KW		cranial trauma; cerebral trauma; spinal cord injury; stroke;	
KW		demyelinating disease; multiple sclerosis; monophasia demyelination;	
KW		encephalomyelitis; multifocal leukoencephalopathy; panencephalitis;	
KW		Marchiafava-Bignami disease; pontine myelinolysis; adrenoleukodystrophy;	
KW		Pelizaeus-Wertheimer disease; spongy degeneration; Alexander's disease;	
KW		Canavan's disease; metachromatic leukodystrophy; viral infection;	
KW		Krabbe's disease.	
XX			
OS		Homo sapiens.	
XX			
FH		Key	
FT		Domain	
FT			Location/Qualifiers
FT			1034..1119
FT			/label= "luminal extracellular domain
FT			/note= "This sequence is specifically claimed"
FT		Peptide	1055..1094
FT			/label= Pept1
FT			/note= "Receptor binding inhibitory peptide. This
FT			sequence is specifically claimed"
FT		Peptide	1064..1088
FT			/label= Pept2
FT			/note= "Receptor binding inhibitory peptide. This
FT			sequence is specifically claimed"
FT		Peptide	1074..1098
FT			/label= Pept3
FT			/note= "Receptor binding inhibitory peptide. This
FT			sequence is specifically claimed"
FT		Peptide	1084..1108
FT			/label= Pept4
FT			/note= "Receptor binding inhibitory peptide. This
FT			sequence is specifically claimed"
FT		Peptide	1095..1119
FT			/label= Pept5
FT			/note= "Receptor binding inhibitory peptide. This
FT			sequence is specifically claimed"
XX			
PD	WO200151520-A2.		
XX			
PD	19-JUL-2001.		
XX			
PF	12-JAN-2001; 2001WO-US01041.		
XX			
PR	12-JAN-2000; 2000US-0175707.		
PR	26-MAY-2000; 2000US-0207356.		
PR	29-SEP-2000; 2000US-0236378.		
XX			
PA	(UYVA) UNIV YALE.		
XX			
PI	Strittmatter SM;		
XX			
DR	WPI; 2001-442138/47.		
DR	N-PSDB; AAS09453.		
XX			
PT			
PT			

nervous system disorders -

Example 1; Page 101-104; 109pp; English.

The sequence is the human Nogo protein, a 250kDa myelin-associated axon growth inhibitor. The invention relates to the use of the nogo receptor CC against them, their nucleic acids, vectors expressing them and antibodies against them, to isolate agents which block nogo receptor mediated axonal growth. The agent is useful for treating a central nervous system disorder which is a result of cranial or cerebral trauma, spinal cord injury, stroke or a demyelinating disease selected from multiple sclerosis, monoplhasis demyelination, encephalomyelitis, multifocal leukoencephalopathy, paraneoplasia, Marchiafava-Bignami disease, pontine myelolysis, adrenoleukodystrophy, Pelizaeus-Merzhafer disease, Sporgy degeneration, Alexander's disease, Canavan's disease, metachromatic leukodystrophy, viral infection and Krabbe's disease.

Sequence 1192 AA;

Query Match Best Local Similarity 7.1%; Score 118; DB 22; Length 1192; Matches 36; Conservative 8; Mismatches 33; Indels 14; Gaps 6

Db 132 NTAAHTPPSPV---AVQSSRPVQOHPAVOKPTPEVVVK---KPTPEPVVQOPAPVAPV 185
115 STVPASPPLSAAAVSPSKLPEDDEFPARPPPAPPASVPSQAEPFVTTPA---PAPRAPDS 171

Oy 186 T-EAAPFAGSGGVMQ---FRYPVGATNPYVR 212
172 TPAAFKRRGGSGVDETFLALP-AASEPVR 201

RESULT 35
AAB82349
ID AAB82349 standard; Protein; 1192 AA.
AC AAB82349;
DT 23-JUL-2001 (first entry)
XX
DE Human NOGO-A protein.
KW NOGO-A; human; chromosome 2p21; neuropathy; spinal injury;
KW brain injury; stroke; neuronal degeneration; Alzheimer's disease;
KW Parkinson's disease; neuromuscular disorder; psychiatric disorder;
KW developmental disorder; neuroprotective; motropic; neuroleptic;
KW hctloParkinsonian; cerebroprotective; neuroleptic; diagnosis;
KW therapy.
OS Homo sapiens.
XX
PN WO200136631-A1.
XX
PD 25-MAY-2001.
XX
PF 14-NOV-2000; 2000WO-GB04345.
XX
PR 15-NOV-1999; 99GB-0026995.
XX
PR 24-JAN-2000; 2000GB-0001550.
XX
FA (SMK) SMITHLINE BEECHAM PLC.
XX
FI Michalovich D, Prinjha R,
XX
DR WP1; 2001-34382/36.
XX
N-PSDB; AAF90324.
XX
FT New polypeptide designated NOGO-C is a splice variant of the human NOGO
FT gene and may be useful in the treatment of neural disorders including
FT Alzheimer's and Parkinson's diseases
XX
XX Disclosure; Page 26-27; 25pp; English.

CC experimentally and clinically without fear of severe infection and/or
 CC contamination.
 XX
 SQ Sequence 1708 AA;

Query Match 7.1%; Score 118; DB 20; Length 1708;
 Best Local Similarity 20.3%; Pred. No. 0.63;
 Matches 79; Conservative 44; Mismatches 129; Indels 138; Gaps 17;

1 MVTAINSONOKPIKRLGIVITTCIAGC--ASKPTNSTSGSGSHRSGGCLA 57
 548 LTAIVELTASPDRECRVLTGNKTRFTTVDGALHLEANGPEQVLSFDASRQSGAGSHS 607
 58 I-----GSOVITDSQGVNRYOVKQDITS-----KIAORYGLN-- 91
 608 LTYELTPAGLQVARISSNGLDCTATFPFGAPSAAPGEVAAPGSAIYRNRFTQNSLTGG 667
 92 -WR-----EIGHI-----NNLNSYTYITGOMLTMSGDLKXREISSGVNTA 134
 668 LMLHREGULGIFPPSPDGHIMEBANPFCGEGTLYTRWST-----SGPSSD 713
 135 HTPEPVAVQSSRPVQOHPAVOKPPTV----- 162
 714 FSP-----EAAAPVLAAPGLPHPTPVSQDIWLPPEKESQVDAASVPAPPEAGLPS 769
 163 VVAKKPTPTPPVQAPVAPVAPTEAPFATSSGVQMRYPVGTNPVRRFGTATVAGS 222
 770 IVTLTPPLPPVAKPPTP-----STRRLIYTPDGA-----KYAAGS 809
 223 TVTNSGMWFSGRDGLINASNAGTVIOADHMDGASIVIQHTNGFVSYHIKDAQVKTG 282
 810 LFESDCNW-----LVNASNPG-----HRPGG---LCH-----AFYQRFPEAYPTE 848
 283 DIVRTGORIASMKNOP-SGAALFEPRISSN 311
 849 FIMREGLAAYTLTPPILIHAVAPDVRVEQN 878

RESULT 38
 AAE18363
 ID AAE18363 standard; Protein, 364 AA.

AC AAE18363;
 XX
 DT 07-MAY-2002 (first entry)
 XX
 DE Streptococcus pyogenes strain SP57 BVH-P1 mature protein.
 XX
 KW BVH-P1 gene; streptococcal infection; pharyngitis; erysipelas; impetigo;
 KW scarlet fever; bacteremia; necrotizing fasciitis; toxic shock; vaccine;
 KW immune response; anti-inflammatory; immunisation; antibacterial.
 XX
 OS Streptococcus pyogenes SP57.
 XX
 PN WO200204495-A2.
 XX
 PD 17-JAN-2002.
 XX
 PF 06-JUL-2001; 2001MO-CA01001.
 XX
 PR 06-JUL-2000; 2000US-216465P.
 XX
 PA (SHIR-) SHIRE BIOCHEM INC.
 XX
 PI Martin D, Hamel J, Brodeur B;
 XX
 DR WPI: 2002-171701/22.
 XX
 DR N-PSDB; AAD29298.
 XX
 PT New Streptococcus pyogenes antigen useful for diagnosing, preventing or
 PT treating streptococcal infection, e.g. pharyngitis, erysipelas and
 PT impetigo, scarlet fever, and invasive diseases

PS Claim 21; Fig 14; 74pp; English.

XX The invention relates to antigens, more particularly an antigen of
 CC Streptococcus pyogenes (also called group A Streptococcus (GAS))
 CC bacterial pathogen. The polypeptides and polynucleotides encoding them
 CC are useful for diagnosing, preventing or treating streptococcal
 CC infection, such as pharyngitis, erysipelas, impetigo, scarlet fever,
 CC invasive diseases (bacteremia, necrotizing fasciitis, toxic shock), and
 CC for eliciting an immune response. The polypeptides may also be used as
 CC immunogens for producing antibodies, or for passive immunisation. DNAs encoding
 CC Streptococcus pyogenes infection, or for passive immunisation. DNAs encoding
 CC polypeptides may also be used to design DNA probes for detecting the
 CC presence of Streptococcus in biological samples suspected of containing
 CC the bacteria. The vaccine composition is useful as a prophylactic or
 CC therapeutic treatment of Streptococcal infection in an individual
 CC susceptible to or infected with streptococcal infection. The present
 CC sequence is Streptococcus pyogenes strain SP57 BVH-P1 mature protein.

SQ Sequence 364 AA;

Query Match 7.0%; Score 117.5; DB 23; Length 364;
 Best Local Similarity 23.5%; Pred. No. 0.094;
 Matches 56; Conservative 22; Mismatches 87; Indels 73; Gaps 8;

73 YOVKQDITVSKIAQRYGLNMRIGHINNLSYTYITGOMLT-----MSGDLKXREIS 126
 24 IYKYGDTLSTIAEMGIDVHVDINHIANIDILFPDTLITANNVQHQAINTLV-QAP 82
 127 ISSGVNTAHTPS-----PVAVQSSRPVQOHPAV---OKPPTPVVYK----- 166
 83 ASPPASVSHVPSSEPLPQASATSQPTVPMAPPATPSDVPPTTPASAKPSSVTASSELTS 142
 167 -----KPTPTPPVQOAPVAPVAPTEAPFATSSGVQMRYPVGTNPVRRFGTATVAGS 222
 143 STNDVSTELSSSESKQPEVPEQAVPTPKAETTEVEPKTIDIS-----EAPTSANRPVNE 197
 214 PGTATVAG-----STVTSNGMWFSGRDGLINASNAG 245
 198 SASSEVSAPAPAPAKETKETSAPAAQKAVADTTSVANSNGLSTA--PNAHYNPMNAG 253

RESULT 39

ABP25889
 ID ABP25889 standard; Protein, 389 AA.

AC ABP25889;
 XX
 DT 02-JUL-2002 (first entry)
 XX
 DE Streptococcus pyogenes strain SP57 BVH-P1 mature protein.
 XX
 KW Streptococcus pyogenes; GAS; group A streptococcus; Streptococcus agalactiae;
 KW group A streptococcus; Streptococcus pyogenes; antibacterial;
 KW anti-inflammatory; infection; vaccine; meningitis; gene therapy.
 XX
 OS Streptococcus pyogenes.
 XX
 PN WO200234771-A2.
 XX
 PD 02-MAY-2002.
 XX
 PF 29-OCT-2001; 2001MO-GB04789.
 XX
 PR 27-OCT-2000; 2000GB-0026333.
 XX
 PR 24-NOV-2000; 2000GB-0028727.
 XX
 PR 07-MAR-2001; 2001GB-0005640.
 XX
 PA (CHIR-) CHIRON SPA
 XX
 PI (GENO-) INST GENOMIC RES.
 XX
 DR Telford J, Masignani V, Margarit Ros YI, Grandi G, Fraser C,
 XX
 PI Tettelein H;

DR MPI, 2002-352536/38.
 DR N-PSDB; ABN66520.
 XX
 XX
 PT New Streptococcus protein for the treatment or prevention of infection
 PT or disease caused by Streptococcus bacteria, such as meningitis, and
 PT for detecting a compound that binds to the protein -
 XX
 XX
 PS Claim 1; Page 3249; 4525dp; English.
 XX
 XX The invention relates to a protein (ABP25413-ABP30895) from group B
 CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (51), given in
 CC the specification. The proteins have antibacterial and anti-inflammatory
 CC activity. (1), nucleic acids encoding (1), ABN6604-ABN71526 and
 CC antibodies that bind (1) are used in the manufacture of medicaments for
 CC the treatment or prevention of infection or disease caused by
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
 CC Nucleic acids encoding (1) are used to detect Streptococcus in a
 CC biological sample. (1) is used to determine whether a compound binds to
 CC (1). A composition comprising (1) or a nucleic acid encoding (1), may be
 CC used as a vaccine or diagnostic composition. The disease caused by
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic
 CC acid encoding (1) may be used to recombinantly produce (1) and may be
 CC used in gene therapy. Antibodies to (1) are used for affinity
 CC chromatography, immunoassays, and distinguishing/identifying
 CC Streptococcus proteins.
 CC
 XX
 XX Sequence 389 AA;
 SQ
 Query Match 7.0%; Score 117.5; DB 23; Length 389;
 Best Local Similarity 23.5%; Pred. No. 0.1;
 Matches 56; Conservative 22; Mismatches 87; Indels 73; Gaps 8
 QY 73 YQVKGQDVSXKIAQRVGLNWEIRIGINNINSSVTYTGQWTL-----WSGDLKVRERS 126
 Db 49 YTVKKGDLSTLAEMGIDVHVLGGINNIHANIIDLFPDITLLANNOHQAINLTV-QAP 107
 QY 127 ISSGVNATHTS---PVAVOSRRPPVQOHPAV---QKTPPPVVVVK----- 166
 Db 108 ASSPASVSHVPSSEPLPOASATSGTVPMPAPATPSDVPTTFPASAKPDSSTVASSELTS 167
 QY 167 -----KTPTPPPVQOAPAPAPVPTTEAPFATSGSGVQOFRPVGATNPVYAR 213
 Db 168 STNDVSTELSSBSQKQPEVPPEQEAIVTPPKAAETTEVEBPKDIS-----EAPTSANRPVNE 222
 QY 214 FGATVAG-----STVTSNGMWFSGRGGDLINASNAG 245
 Db 223 SASSEVSSAAAPQAPEKEEISAPAPQAQAVADTTVSATISNGLSLA--PNAHYNPNNAG 278
 RESULT 40
 AAE18359
 ID AAE18359 standard; Protein; 389 AA.
 XX
 XX AAE18359;
 AC
 XX
 DT 07-MAY-2002 (first entry)
 XX
 DB Streptococcus pyogenes strain SP157 BVH-P1 protein.
 XX
 XX BVH-P1 gene; streptococcal infection; pharyngitis; erysipelas; impetigo;
 KW scarlet fever; bacteraemia; necrotising fasciitis; toxic shock; vaccine;
 XX immune response; anti-inflammatory; immunisation; antibacterial.
 OS Streptococcus pyogenes SP157.
 XX
 FH Key Location/Qualifiers
 FH Peptide 1..25
 FT /label= Signal_peptide
 FT 26..389
 FT Protein
 FT /note= "Mature_BVH_P1_protein"
 XX
 WN0200204495-A2.

Query Match	Best Local Similarity	Score	DB 23;	Length 389;
Matches	56;	Conservative	22;	Mismatches 87; Indels 73; Gaps 8;
06-JUL-2001; 2001WO-CA01001.	06-JUL-2001; 2000US-216465P.	06-JUL-2000; 2000US-216465P.	(SHIR-) SHIRE BIOCHEM INC.	Martin D, Hamel J, Brodeur B;
WP1; 2002-171701/22.	N-PSDB; AAD29294.			
New Streptococcus pyogenes antigen useful for diagnosing, preventing or treating streptococcal infection, e.g. pharyngitis, erysipelas and impetigo, scarlet fever, and invasive diseases				
Claim 21; Fig 6; 74pp; English.				
The invention relates to antigens, more particularly an antigen of Streptococcus pyogenes (also called group A Streptococcus (GAS)) bacterial pathogen. The polypeptides and polynucleotides encoding them are useful for diagnosing, preventing or treating streptococcal infection, such as pharyngitis, erysipelas, impetigo, scarlet fever, invasive diseases (bacteremia, necrotising fasciitis, toxic shock), and for eliciting an immune response. The polypeptides may also be used as immunogens for producing antibodies for the diagnosis and treatment of Streptococcus infection, or for passive immunisation. DNAs encoding polypeptides may also be used to design DNA probes for detecting the presence of Streptococcus in biological samples suspected of containing the bacteria. The vaccine composition is useful as a prophylactic or therapeutic treatment of Streptococcal infection in an individual susceptible to or infected with streptococcal infection. The present sequence is Streptococcus pyogenes strain SPY5 BVH-P1 protein.				
Sequence	389 AA;			
73 YQVQSGTSTKIAQRYGLNMRREIHNINNSYITTYGQMTL-----WGDGLKVRERS	126			
49 YTVRYGGTLSTIAEMGIDVHVLIDINHINAIIDLIPTTLITLVANYNOHGQATNLTV-QAP	107			
127 ISSGVNFAHTPS-----PVAOSSRPVQOHPAV---QKPTPPVVVK-----	166			
108 ASSPASVSHVSSSEPLDQASATSGPYTPMARPAPPSDVFPTTFPFAKDDSSVTASSELTS	167			
167 -----KPTTPPVVQOPAVPAPVTEAPFATGSSGGWQFRYPVGATNPVVR	213			
168 STNDVSTELSGESQKQEPVPEQAVPTPKAAETTEVEPKTIDS-----EAPTSARPVNE	222			
214 FGLTIVG-----STYTSNQMWFSGRGDGLINSNAG	245			
223 SASSEVSAPAPQAPAEKETSTAPAPQAKAVADTTVAISNLSLSTA--PNAHYNPWNA	278			

Search completed: July 8, 2003, 11:05:36
Job time : 74 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 8, 2003, 11:04:38 ; Search time 26 Seconds
(without alignments)
364.392 Million cell updates/sec

Title: US-10-018-706-2
Perfect score: 1670
Sequence: 1 MVTITAINSGNQKPIKRLGL.....LFEFRISRGVYDPLTVLK 332

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	209	12.5	345	4	US-09-405-728-5
2	151.5	9.1	478	2	US-08-456-6708-40
3	146.5	8.8	484	1	US-08-127-499A-26
4	146.5	8.8	484	1	US-08-482-847-26
5	130.5	7.8	437	2	US-08-737-716-2
6	123.5	7.4	666	2	US-08-737-716-14
7	121.5	7.3	232	2	US-08-456-6708-39
8	119.5	7.2	1426	4	US-09-136-574A-43
9	118	7.1	216	4	US-08-737-716-4
10	118	7.1	1708	4	US-09-462-606-2
11	111	6.6	330	4	US-09-134-001C-4002
12	110.5	6.6	1751	4	US-09-136-574A-44
13	109.5	6.6	480	3	US-09-189-035-5
14	109.5	6.6	480	4	US-09-382-086-5
15	109.5	6.6	616	4	US-09-136-574A-47
16	109	6.5	907	3	US-08-783-774-2
17	109	6.5	907	3	US-09-338-599A-1
18	109	6.5	907	5	PCT-US95-04611A-19
19	108.5	6.5	2285	4	US-09-308-375-2
20	108	6.5	265	3	US-08-619-812-9
21	108	6.5	279	3	US-08-619-812-4
22	107.5	6.4	671	2	US-08-737-716-13
23	107	6.4	490	4	US-09-109-841-2
24	106.5	6.4	878	4	US-09-556-706B-2
25	103.5	6.2	250	1	US-08-155-171B-15
26	103.5	6.2	250	2	US-08-435-998-15
27	103	6.2	214	1	US-08-217-327-4

28	102	6.1	447	4	US-09-120-927-2	Sequence 2, Appli
29	101.5	6.1	521	1	US-08-276-213-3	Sequence 3, Appli
30	101.5	6.1	1162	2	US-08-728-323A-2	Sequence 2, Appli
31	101.5	6.1	1162	4	US-09-298-568-2	Sequence 2, Appli
32	100.5	6.0	1130	4	US-09-442-100-4	Sequence 4, Appli
33	98.5	5.9	268	1	US-08-431-387-4	Sequence 4, Appli
34	98.5	5.9	750	4	US-09-165-239A-4	Sequence 4, Appli
35	98.5	5.9	1702	4	US-08-296-791-5	Sequence 5, Appli
36	98.5	5.9	1702	5	PCT-US95-10661A-5	Sequence 5, Appli
37	98	5.9	1693	3	US-08-478-507-7	Sequence 7, Appli
38	98	5.9	1693	4	US-09-128-275A-7	Sequence 7, Appli
39	98	5.9	1693	4	US-09-553-427-7	Sequence 7, Appli
40	97	5.8	2972	4	US-09-579-181-2	Sequence 2, Appli
41	97	5.8	3118	4	US-09-579-181-1	Sequence 1, Appli
42	96	5.7	442	2	US-08-821-355A-5	Sequence 5, Appli
43	96	5.7	442	2	US-09-003-687A-5	Sequence 5, Appli
44	96	5.7	442	2	US-09-136-605-5	Sequence 5, Appli
45	96	5.7	596	2	US-08-821-355A-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1
US-09-405-728-5
Sequence 5, Application US/09405728
Patent No. 6391316
GENERAL INFORMATION:
APPLICANT: Potter, Andrew A.
APPLICANT: Rieux, Clement
APPLICANT: Schriyers, Anthony B.
TITLE OF INVENTION: CLONING AND EXPRESSION OF HAEMOPHILUS SOMMUS
TITLE OF INVENTION: TRANSFERRIN-BINDING PROTEINS
FILE REFERENCE: 9000-0049.20
CURRENT APPLICATION NUMBER: US/09/405,728
CURRENT FILING DATE: 1999-09-24
EARLIER APPLICATION NUMBER: US 09/267,749
EARLIER FILING DATE: 1999-03-10
NUMBER OF SEQ ID NOS: 5
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 5
LENGTH: 345
TYPE: PRT
ORGANISM: Haemophilus sommus
US-09-405-728-5

Query Match 12.5%; Score 209; DB 4; Length 345;
Best Local Similarity 24.8%; Pred. No. 1.5e-11;
Matches 63; Conservative 49; Mismatches 106; Indels 36; Gaps 6;
QY 73 YQKQGDVTSKIAQRYGELNREIGHNINNSYTYTGQWLTMSG---DLKREKISS 129
DB 120 YKVRKGDVTFLLIYISGMDIKELATLNNMSEPHLSIGQVLKANNIPDSNM1PTQTIME 179
QY 130 GVTWATPSPVAQSSRPVQHPAVQKPPPVVVVKKPPPTPVVQOPAPVAPVTEAP 189
DB 180 SEYTOVTNVE-TNNANKPTQKPKVATPTHTSMPIK---TEPATNTIMWP----- 229
QY 190 FATGSSGWMQFRYPVQATNPFVERFGTATVAGSTVTSNGMPSGRODGLINASNAQTIVQ 249
DB 230 -----TNGKIIQGFSSADG-----NKGIDISGSGQAVNAAGRVVY 268
QY 250 ADINMGAS--IYIOHTNGFVSYSYIHKDAQVKTGPTVRTGORTASMKNOPSGAALFEFR 307
DB 269 AGDALRKYGKLLTIKNDSTLSAYAHNESILVQKQGEVKAQGOIARKGSSGTITIKLHFE 328
QY 308 ISRNGVYVDPLTVL 321
DB 329 IRYKQGSVDPMRWL 342
RESULT 2
US-08-456-670B-40

Wed Jul 9 10:00:59 2003

us-10-018-706-2.ral

Page 2

Sequence 40, Application US/08456670B
Patent No. 5932415
GENERAL INFORMATION:
APPLICANT: SCHUBERT, PETER
APPLICANT: NEUMANN, Siegfried
APPLICANT: PAMELZIK, MARTINA
APPLICANT: LINKEWELER, WINFRIED
APPLICANT: BURGER, CHRISTA
APPLICANT: HOFMANN, GOTTFRIED
APPLICANT: HUBERT, ANDREAS
APPLICANT: GORBEL, STEFAN
APPLICANT: KOHLER, STEFAN
TITLE OF INVENTION: PROCESSES AND AGENTS FOR DETECTING
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: MILLER, WHITE, ZELANO & BRANNIGAN, P.C.
STREET: 2200 CLARENDON BLVD., SUITE 1400
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: US
ZIP: 22201
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/456,670B
CLASSIFICATION: 435
FILING DATE: 01-JUN-1995
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/412,227
FILING DATE: 27-MAR-1995
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/075,248
FILING DATE: 11-JUN-1993
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: DE 4239567.4
FILING DATE: 25-NOV-1992
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: DE 421911.4
FILING DATE: 11-JUN-1992
ATTORNEY/AGENT INFORMATION:
NAME: HAMLET-KING, DIANE
REGISTRATION NUMBER: 33,302
REFERENCE/DOCKET NUMBER: MERCK 1694D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-243-6333
TELEFAX: 703-243-6410
INFORMATION FOR SEQ ID NO: 40:
SEQUENCE CHARACTERISTICS:
LENGTH: 478 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
ORIGINAL SOURCE:
ORGANISM: Listeria monocytogenes
STRAIN: EGD
US-08-456-670B-40

Query Match 9.1%; Score 151.5; DB 2; Length 478;
Best Local Similarity 21.3%; Pred. No. 5.7e-06;
Matches 83; Conservative 46; Mismatches 117; Indels 143; Gaps 15;

QY 25 ITTCLILGCAKSPYNTSGSGSHRTSGGGLA--IGSQVITD----- 65
DB 102 IITSIKGG--TKYVETTESNGWTKITVNDGKTFVNGKYLDDKAVSTPAAPQEVKKEET 159
QY 66 --SGGVP-----NRQYKQGDPTVSKIAQRYG 89
DB 160 TTQGAARVAETKEVKTQTTQATTPAPKVAETKETPTVIDQNTTAAVSGDTTIALSVKYG 219

QY 90 LNMREIGHINNINSYTYITGQMLTMSGDLKVRERSISSGVNTAHTSPVAVOSRPPV 149
DB 220 VSVQDIMSNNLSS--SIYVGQKAI-----KQANTATPPAEVKT----- 259
QY 150 QCHPAVQKPTPPV-----VKKPTPPVQOPAPVAPVTE-----APPATG 193
DB 260 -EAPAEKQAPVVENNTNTNTATTEKEETATQ--QOTAPKAP--TEAPKAPAPSTNT 313
QY 194 SSGVMQFRYPVATNPVVARFGTATVAGSTVNSGMWFGSGDGLINASNGVTOADHN 253
DB 314 NAKNTNTNTNTNTNTPSKNTNTNTNTNTNTNTNSNANAGSSNNNSASAIIAEQKH 373
QY 254 MD-----GASIVIOHTNGF--VASYTHIKDAQVKTGDTVRTG 288
DB 374 LKAVSWGNGDPTEDCSGYTKVPKAGISLPRTSGQVASTTRISGSQKKPDLY--- 430
QY 289 ORIASMKQPSGALFEF--RISRRGVVY 315
DB 431 -----FDDYGGSGISHVGIYV 445

RESULT 3
US-08-127-499A-26
Sequence 26, Application US/08127499A
Patent No. 5510264
GENERAL INFORMATION:
APPLICANT: VAN ALSTYNE, Diane
APPLICANT: SHARMA, Lawrence Rajendra
TITLE OF INVENTION: ANTIBODIES WHICH BIND MENINGITIS RELATED
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/127,499A
FILING DATE: 28-SEP-1993
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 51916/102/INBI
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 484 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
US-08-127-499A-26

Query Match 8.8%; Score 146.5; DB 1; Length 484;
Best Local Similarity 20.6%; Pred. No. 1.7e-05;
Matches 81; Conservative 46; Mismatches 121; Indels 145; Gaps 14;

QY 25 ITTCLILGCAKSPYNTSGSGSHRTSGGGLA--IGSQVITD----- 65
DB 102 IITSIKGG--TKYVETTESNGWTKITVNDGKTFVNGKYLDDKAVSTPAAPQEVKKEET 159
QY 66 --SGGVP-----NRQYKQGDPTVSKIAQRYG 89

```

Db      160  TTQQAAPAEETKEVEKQTTQATTAPPAKVETKEPEVQDNQNTTHAVXSGDTIMALSXYG  219
Qy      90  LNNREIGHNINNLSSTIYTGOMLMSDDLKYBRERSISSGWNTHPSPYAVQSSRPV  149
Db      220  VASQDDIMSNANSSSS-SIYQOKLAI-----KQTNATATPPAEVKT-----  259
Qy      150  QQHFAVOKPTPPV-----VKKPTTPPVQQAPAPAPVPEAEPAATSSGWMQ  199
Db      260  -EAPPAEKQAPVYKENTNTNTATTEKETATQ--QQTAPAPAPAEAKPAPASTNTNA  315
Qy      200  FRAPVGA-----TNPVVRREGTAVASITYSNGMFFSGDDGLIASAGVVIQ  249
Db      316  NKNTNTNTNTNTNTNTNTNTNTPSKNTVNTSNNTNTNSNTNANQSSNNNSSSAIIAE  375
Qy      250  ADHNM-----GASIVIOHTNGF-VSSYIHKDAQVKTGDT  284
Db      376  AQKHLKAVSWGNGEPPTFDSCGYTKYVPAKAGISLPTSGAQVASTRISSESQAKGDL  435
Qy      285  VRTGRIASMKQNPSCAALFEF--RISRGYVV  315
Db      436  V-----FPGYSGSGSHVGIYV  451

```

RESULT 4
US-08-482-047-26
; Sequence 26, Application US/08482847
; Patent No. 5556757
; GENERAL INFORMATION:

APPLICANT: VAN ALSTYNE, Diane
 APPLICANT: SHARMA, Lawrence Rajendra
 TITLE OF INVENTION: PEPTIDES REPRESENTING EPITOPIC SITES FOR
 TITLE OF INVENTION: BACTERIAL AND VIRAL MENINGITIS CAUSING AGENTS AND THEIR
 TITLE OF INVENTION: CNS CARRIER, ANTIBODIES THEREO, AND USES THEREOF
 NUMBER OF SEQUENCES: 40
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Foley & Lardner
 STREET: 3000 K Street, N.W., Suite 500
 CITY: Washington
 STATE: D.C.
 COUNTRY: USA
 ZIP: 20007-5109
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/482,847
 FILING DATE: 07-JUN-1995
 CLASSIFICATION: 514
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/127,499
 FILING DATE: 28-SEP-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: BENT, Stephen A.
 REGISTRATION NUMBER: 29,768
 REFERENCE/DOCKET NUMBER: 51916/104/INBI
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202) 672-5300
 TELEFAX: (202) 672-5399
 TELEX: 904136
 INFORMATION FOR SEQ ID NO: 26:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 484 amino acids
 TYPE: amino acid
 STRANDEDNESS:
 TOPOLOGY: unknown
 US-08-482-847-26

Query Match	8.8%	Score 146.5;	DB 1;	Length 484;
Best Local Similarity	20.6%	Pred. No. 1.7e-05;		
Matches 81; Conservative	46;	Mismatches 121;	Indels 145;	Gaps 14;

```

OY      25 IITCIIAGASCRTYVSTSGSHSTSSGGIA--IGSOVTD-----65
Db      102 IITISIKGG--TKYVETTESNGMHKITYNDGKTGFVNGSKYILTDAAVSTPAAPIDOEYKKT 159
OY      66 --SQGVV-----NRYOXQGDVTSKIAORYG 89
Db      160 TTQQAAPAAETKEVYKQTQATTAPAPKVAETKEPVVDQNAATHVAHSGDITWALSXYG 219
OY      90 LMRREIGHINNLNSSTITYTGOMLLMSGDLKVRERSISSGVNTAHPSPVAVOSSRPV 149
Db      220 VSVODLMSNNLSSS--SIYVGQKAI-----KQAPANTAPAEAEVKT-----259
OY      150 QQHAPAYOKRPVY-----VKKPPTPVPVQOAPAPVAPVTEAPATGSSGVWQ 199
Db      260 -EAPAEKQAPVYKENTNTNTATTETKEKETATQ--QDTAKATTEAKKAPAPASTTNA 315
OY      200 FRYPVGA-----TNPVRRFGTAVGASTVTSNGMFSGRDGLINASNAQTVIQ 249
Db      316 NKNTNTYNTNTNTNTNTNTNTPSKNTNTNSNTNTNSNTNANQSSNNNSSSASALIAE 375
OY      250 ADHNMD-----GASIVYQHTNGF--VSYIHIKDAQVKTGDT 284
Db      376 AQCHLKAYSWGNGEPTTDCSGYTKYVFAVAGISLPTSGAQAYSTRTISESQAKPDL 435
OY      285 VRTGRIASMKONPGSALFEF--RISNGVYV 315
Db      436 V-----FEDYGGGISHVGIIV 451

```

RESULT 5
US-08-737-716-2
; Sequence 2, Application US/08737716
; Patent No. 5955258

GENERAL INFORMATION:
 APPLICANT: GIBBE, BUIST
 APPLICANT: Gerard VENEMA
 APPLICANT: Jan KOK
 APPLICANT: Adrianus Marinus LEDBOER
 TITLE OF INVENTION: Process for the lysis of a culture of lactic
 TITLE OF INVENTION: acid bacteria by means of a lysin, and uses of the resulting
 TITLE OF INVENTION: lysed culture.
 NUMBER OF SEQUENCES: 16
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Pillsbury Madison & Sutro, L.L.P.
 STREET: 1100 New York Avenue, N.W.
 CITY: Washington
 STATE: D.C.
 COUNTRY: U.S.A.
 ZIP: 20005-3918
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Microsoft Word
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/737,716
 FILING DATE: 22-APR-1997
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: PCT/NL95/00170
 FILING DATE: 12-MAY-1995
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: EP 94201353.3
 FILING DATE: 12-MAY-1994
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 437 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-737-716-2

Query Match 7.8%; Score 130.5; DB 2; Length 437;

Wed Jul 9 10:00:59 2003

us-10-018-706-2.ra1

Page 4

Best Local Similarity 24.1%; Pred. No. 0.00046;
Matches 69; Conservative 39; Mismatches 119; Indels 59; Gaps 12;

Qy 26 TCCILAGCAKSKPTNST-----SGSGSHRTSGSGGLAIGQVITDSQGVN- 71
Db 184 TALGRTAIDPSIGASINRIISQYLTFRFGASAGNINSGS---STTTINNNSGINS 240
Qy 72 ---RVOYQCDTVSKIAORGLMREIGHINNNSYTYTQOMTLMSGDLKVRERSIS 128
Db 241 SSTIYVSGDITLWGISCFYGISVAQIOSANLNKST-IYIQQLVLVGS---ASSTNS 295
Qy 129 SGVNTATPSPVAVOSSRPVQOHPAVOKPTPPVVVKKPTPT--FPVQOPAPVAPPT 186
Db 296 GGSNNSS-----ASTPTTSTVTPA--KPTSGTTVKKSGDITLMSLVKTSIAQLKS 345
Qy 187 EAPATGSSGVQWQ---FRYPVGNTPVVRFGTATVAGSTVTSNGMFGSGRGGDLINASN 243
Db 346 WNLHSDTYITIGNLIVSQAASANP-----STGSGSTATNNNSSTSSNASHKV 398
Qy 244 AGTVIQADHNMDSASIVIOHTNGFVSSYIHKDAQVTKGDTVTRGQ 289
Db 399 KQDTLWGLSQKSGSPI-----ASIKAWNHL-----SDITILLGQ 432

RESULT 6

US-08-737-716-14
; Sequence 14 Application US/08737716
; Patent No. 595258
GENERAL INFORMATION:
APPLICANT: Gilbe BUIST
APPLICANT: Gerard VENEMA
APPLICANT: Van KOK
TITLE OF INVENTION: Adrianus Martinus LEDERBOER
TITLE OF INVENTION: Process for the lysis of a culture of lactic
TITLE OF INVENTION: acid bacteria by means of a lysin, and uses of the resulting
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pillsbury Madison & Sutro, L.L.P.
STREET: 1100 New York Avenue, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005-3918
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Microsoft Word
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/737,716
FILING DATE: 22-APR-1997
CLASSIFICATION: 435
PRIOR APPLICATION NUMBER:
APPLICATION NUMBER: PCT/NL95/00170
FILING DATE: 12-MAY-1995
APPLICATION NUMBER: EP 94201353.3
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 666 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Enterococcus hirae
IMMEDIATE SOURCE:
CLONE: Fig-5a (E. hirae)
US-08-737-716-14

Query Match 7.4%; Score 123.5; DB 2; Length 666;
Best Local Similarity 31.6%; Pred. No. 0.0038;

Matches 36; Conservative 21; Mismatches 40; Indels 17; Gaps 4;

Qy 31 AGCAKPTNTVSG---SGSHRTSGSGGLAIGQVITDSQGVNRYQKQSDTVSKIAOR 87
Db 304 AGSSTVNTGNASSGNTSGNTSGTSGTQATGA-----KTVKSGDSVWIKAND 352
Qy 88 YGLMWRIGHINNNSYTYTQOMTLMSGDLKVRERSISGVNTAHTSPVA 141
Db 353 HGISMNOLIMNNIKNNF-VYPQQLVYSKSSSA--SGSTNSTGTNTSSNTA 403

RESULT 7

US-08-456-6708-39
; Sequence 39 Application US/084566708
; Patent No. 5932415
GENERAL INFORMATION:
APPLICANT: SCHUBERT, PETER
APPLICANT: NEWMANN, SIEGFRIED
APPLICANT: FAMELZIK, MARTINA
APPLICANT: LINKWIELER, WINFRIED
APPLICANT: BORGER, CHRISTA
APPLICANT: HOFMANN, GOTTFRIED
APPLICANT: BOBERT, ANDREAS
APPLICANT: KOHLER, WERNER
TITLE OF INVENTION: PROCESSES AND AGENTS FOR DETECTING
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: MILLER, WHITE, ZELANO & BRANIGAN, P.C.
STREET: 2200 CLARENDON BLVD., SUITE 1400
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: US
ZIP: 22201
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/456,6708
FILING DATE: 01-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION NUMBER:
APPLICATION NUMBER: US 08/412,227
FILING DATE: 27-MAR-1995
APPLICATION NUMBER: US 08/075,248
FILING DATE: 11-JUN-1993
APPLICATION NUMBER: DE 4239567.4
FILING DATE: 25-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DE 4219111.4
FILING DATE: 11-JUN-1992
ATTORNEY/AGENT INFORMATION:
NAME: HAMLET-KING, DIANA
REGISTRATION NUMBER: 33,302
REFERENCE/DOCKET NUMBER: MERCK 1694D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-243-6333
TELEFAX: 703-243-6410
INFORMATION FOR SEQ ID NO: 39:
SEQUENCE CHARACTERISTICS:
LENGTH: 232 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
ORIGINAL SOURCE:
ORGANISM: Listeria monocytogenes
STRAIN: EGD

QY 91 NMREIGHNINNSYTYTGQWLTMSGDLKVRERSISSGVNTAHTSPVAVOSSRPPVQ 150
Db 42 SVAOIOSANNLKST-IITYGQGLVLTGS-----ASTTPTS 88
QY 151 QHPAVOKPTPPVAVVKKPTPT--PPVVOGPAVPAPVPEAFATGSSGWMQ---FRYPVG 205
Db 89 VTPA--KPTSGTTVKVGSGDTLMALSVKTKSIQKXSNHLSBDITYIGQNLVSGSA 146
QY 206 ATPVVRFGTATVATGSGTVMFSGRDDLINASAGTIVQADHNMGASIVIOHTN 265
Db 147 ASNP-----STGSGSTRINNSNSTSSNSASHKVKGDTLWGLSKGSP1----- 193
QY 266 GFVSSVYIHKDAQVTKGTVRTGQ 289
Db 194 ASIKAMNH-----SDTILIGQ 211

RESULT 10

US-09-462-606-2
Sequence 2, Application US/09462606
Patent No. 6432408
GENERAL INFORMATION:
APPLICANT: MENG, XIANG-JIN
APPLICANT: Emerson, Suzanne U.
TITLE OF INVENTION: A SWINE HEPATITIS E VIRUS AND USES THEREOF
FILE REFERENCE: 20264267U1
CURRENT FILING DATE: 2000-06-12
CURRENT APPLICATION NUMBER: US/09/462,606
PRIOR FILING DATE: 1997-07-18
PRIOR APPLICATION NUMBER: US 60/053069
PRIOR FILING DATE: 1998-07-17
NUMBER OF SEQ ID NOS: 65
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 2
LENGTH: 1708
TYPE: PRT
ORGANISM: Hepatitis E virus
US-09-462-606-2

Query Match 7.1%, Score 118, DB 4, Length 1708,
Best Local Similarity 20.3%, Pred. No. 0.048,
Matches 79; Conservative 44; Mismatches 129; Indels 138; Gaps 17;

QY 1 MVTIAINSONOKPIKELIGLIFVITTCILAGC---ASKPTYNSTSGSGSHRTSGSGGLA 57
Db 548 LTAIVELTASPDRLCETVIGNKTFTTVVDGAHLEANGPEQYVLSFDASRQMGAGSHS 607
QY 58 I-----GSGVITDSOGVNRXOVKQDVTVS-----KIAORYGLN-- 91
Db 608 LTYELTPAGLQVRISSNGLDCTATFPFGAPSAAPGEVAFAFCALYRNNFTQRHSLTGG 667
QY 92 -WR-----EIGHI-----NNINSYTYTGQWLTMSGDLKVRERSISSGVNTA 134
Db 668 LMAHPEGLIGLFPFSSGHIMEPAPNFCGGLTYTRTWSF-----SGFSSD 713
QY 135 HTPSPVAVOSSRPPVVOQHPAVOKPTPPV----- 162
Db 714 FSPF-----EAAAPVLAAPGLPHPTPVSIDVILPPSKESQVDASVPAPAPALPSS 769
QY 163 VVVKKPTPTPPVVOQAPVAPVPTPEAFATGSSGWMQFRPVATPVVRRFGTATVAGS 222
Db 770 IVTLPLPPLPVKRPFP--P-----STRRLTYTPDGA-----KYVAGS 809
QY 223 TVTNSGWMFSGRDDLINASAGTIVQADHNMGASIVIOHTNGFVSSYIHKDAQVTKG 282
Db 810 LPESSDCWM-----LVNASNPG-----HRPGSG-----AFYQRPPEAPYPT 848
QY 283 DTVRTGORIASMKQNP-SGAALPEPRISRN 311
Db 849 FVMEBGLAAYTLTPRPIIHAADVADRYEON 878

RESULT 11

US-09-134-001C-4002
Sequence 4002 Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Donnette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
FILE REFERENCE: GTC-007
CURRENT FILING DATE: 1998-08-13
CURRENT APPLICATION NUMBER: US/09/134,001C
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 4002
LENGTH: 330
TYPE: PRT
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4002

Query Match 6.6%, Score 111, DB 4, Length 330,
Best Local Similarity 31.0%, Pred. No. 0.02,
Matches 36; Conservative 17; Mismatches 41; Indels 22; Gaps 4;

QY 15 IKRGLIRGVITTCILAGCASKPTYNSTSGSGSHRTSGSGGLAIGSVITDSOGVNRVQ 74
Db 54 IAKLSKINGLISNLIFFPOVUK-----VSGSSRAITSNGTV-----YT 93
QY 75 VKQGTVAKIQORYGLMREIGHNINNSYTYTGQWLTMSGDLKVRERSISSG 130
Db 94 VKAGDSLSIAKYGTYQKIMQNLN-NYLIFPGQKLV-SGKATSSRAVASG 147

RESULT 12

US-09-136-574A-44
Sequence 44, Application US/09136574A
Patent No. 6294366
GENERAL INFORMATION:

APPLICANT: Farrington, Graham K.
Anderson, Paige
Gibbs, Moreland
Bergquist, Peter
Daniels, Roy
Morgan, Hugh W.
Williams, Diane P.

TITLE OF INVENTION: Compositions and Methods for
Treating Cellulose Containing Fabrics Using Truncated
Cellulase Enzyme Compositions

NUMBER OF SEQUENCES: 49
CORRESPONDENCE ADDRESS:

ADDRESSEE: Howson and Howson
CITY: Spring House
STATE: PA
COUNTRY: USA

ZIP: 15477
COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/136,574A
FILING DATE: 19-Aug-1998

CLASSIFICATION DATA:
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/932,571
FILING DATE: September 19, 1997

ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E.

REGISTRATION NUMBER: 31,215
 REFERENCE/DOCKET NUMBER: 1997US001/CIP
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 215-540-9200
 TELEFAX: 215-540-5818
 TELEX: <Unknown>
 INFORMATION FOR SEQ ID NO: 44:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1751 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 SEQUENCE DESCRIPTION: SEQ ID NO: 44:
 US-09-136-574A-44

Query Match 6.6%; Score 110.5; DB 4; Length 1751;
 Best Local Similarity 23.3%; Pred. No. 0.25;
 Matches 74; Conservative 34; Mismatches 99; Indels 111; Gaps 18;

51 SGGGLAIGS-----QVITDSQGVNRYOKQDVTSKIAQRYGLMREIGHINL 101
 1124 SGACVLAAGSTKIRLSIQGSGSYNOSNDYSVRS-----ATGTEKKTGYIDG- 1175
 102 NSSYTIYTGWLTMSGDLKVRERSISSGVNTAH--TPSPVAVOSRPPVQGHPAVOKPT 159
 1176 -----AIWVG-----RPSRGTPAGGVPTPAFTPTSTP-----TPT 1208
 160 PPVVVVKPTPTTPVVOQPAVAPVTE-----APFATGSGVMOFRYPV 204
 1209 PTTT-----PTPTPTVYTPPT-TPAVVPDVKISIDTSRGRKISPIYGANQDIO----- 1258
 205 GATNPVVRFGTATVAGSTVSNQMWFSGRDGLINASNAGT-----VIQ 249
 1259 GVVHP-ARRLG-----GNRLTGYN-WEN-----NNSNAGSDWYHSSDDYMCYIMGITG 1304
 250 ADHNMAGSIVIQH-----TNGFVSYIHIKDAQVKTGD--TVRTGQ-----RIAMKNQOP 298
 1305 NDKVVPAAVYSKPEHQSISKQAVSAILQLQWGVYAKQNGTVSSERFAPSRMAEVVFKK 1364
 299 SGALFFERISRNQVYVD 316
 1365 DGLSLQPDVNDVYVMD 1382

RESULT 13
 US-09-189-035-5
 Sequence 5, Application US/09189035
 Patent No. 6020165
 GENERAL INFORMATION:
 APPLICANT: Yue, Henry
 APPLICANT: Corley, Neil C.
 APPLICANT: Guegler, Karl J.
 APPLICANT: Baughn, Mariah R.
 TITLE OF INVENTION: CYTOKINE SIGNAL REGULATORS
 FILE REFERENCE: PF-0638 US
 CURRENT APPLICATION NUMBER: US/09/189, 035
 CURRENT FILING DATE: 1998-11-10
 NUMBER OF SEQ ID NOS: 6
 SOFTWARE: PERL Program
 SEQ ID NO 5
 LENGTH: 480
 TYPE: PRT
 ORGANISM: Homo sapiens
 FEATURE: -
 OTHER INFORMATION: 92245671
 US-09-189-035-5

Query Match 6.6%; Score 109.5; DB 3; Length 480;
 Best Local Similarity 22.8%; Pred. No. 0.048;
 Matches 53; Conservative 19; Mismatches 79; Indels 81; Gaps 8;
 9 SONQPKIRLGLIFGVTTCILAGCASKPTNSTSGSHRTSGSGGLAIGSQVITDSOG 68

231 SLNQRPTHSG-----SSGSGSRENSSSSIGIPVAPFS-- 267
 69 VPNRYQKQDVTYSKIAQRYGLMREIGHINNLSSTIYTGWLTMSGDLKVRERSIS 128
 268 -PPTIGPAPG---SAPGSQYGTWTRQISRHNSTTS-----STS 301
 129 SGVNTAHTPSPVAVOSRPPVQGH-----AVQKPTPPV----- 163
 302 SG-GYRRTPSVTQAQFAQPHVNGGPLYSONSISIAFPFPMQPLTPQIPLTGFVARVOEN 360
 164 VVKKPTPTP-----VVOQPAVAPVTEAPFATGSSGVMOFRYPVGATNP 209
 361 IADSPTPPPPPDDIPIMFDDSPPPPPPVVDYEDEAAVQVNDPYADGDP 412

RESULT 14
 US-09-382-086-5
 Sequence 5, Application US/09382086
 Patent No. 6201106
 GENERAL INFORMATION:
 APPLICANT: Yue, Henry
 APPLICANT: Corley, Neil C.
 APPLICANT: Guegler, Karl J.
 APPLICANT: Baughn, Mariah R.
 TITLE OF INVENTION: CYTOKINE SIGNAL REGULATORS
 FILE REFERENCE: PF-0638 US
 CURRENT APPLICATION NUMBER: US/09/382, 086
 CURRENT FILING DATE: 1999-08-24
 EARLIER APPLICATION NUMBER: 09/189, 035
 EARLIER FILING DATE: 1998-11-10
 NUMBER OF SEQ ID NOS: 6
 SOFTWARE: PERL Program
 SEQ ID NO 5
 LENGTH: 480
 TYPE: PRT
 ORGANISM: Homo sapiens
 FEATURE: -
 OTHER INFORMATION: 92245671
 US-09-382-086-5

Query Match 6.6%; Score 109.5; DB 4; Length 480;
 Best Local Similarity 22.8%; Pred. No. 0.048;
 Matches 53; Conservative 19; Mismatches 79; Indels 81; Gaps 8;

9 SONQPKIRLGLIFGVTTCILAGCASKPTNSTSGSHRTSGSGGLAIGSQVITDSOG 68
 231 SLNQRPTHSG-----SSGSGSRENSSSSIGIPVAPFS-- 267
 69 VPNRYQKQDVTYSKIAQRYGLMREIGHINNLSSTIYTGWLTMSGDLKVRERSIS 128
 268 -PPTIGPAPG---SAPGSQYGTWTRQISRHNSTTS-----STS 301
 129 SGVNTAHTPSPVAVOSRPPVQGH-----AVQKPTPPV----- 163
 302 SG-GYRRTPSVTQAQFAQPHVNGGPLYSONSISIAFPFPMQPLTPQIPLTGFVARVOEN 360
 164 VVKKPTPTP-----VVOQPAVAPVTEAPFATGSSGVMOFRYPVGATNP 209
 361 IADSPTPPPPPDDIPIMFDDSPPPPPPVVDYEDEAAVQVNDPYADGDP 412

RESULT 15
 US-09-136-574A-47
 Sequence 47, Application US/09136574A
 Patent No. 6294366
 GENERAL INFORMATION:
 APPLICANT: Farrington, Graham K.
 APPLICANT: Anderson, Paige
 APPLICANT: Gibbs, Moreland
 APPLICANT: Bergquist, Peter
 APPLICANT: Daniels, Roy
 APPLICANT: Morgan, Hugh W.

Wed Jul 9 10:00:59 2003

us-10-018-706-2.ra1

Page 8

Williams, Diane P.
TITLE OF INVENTION: Compositions and Methods for
Treating Cellulose Containing Fabrics Using Truncated
Cellulase Enzyme Compositions
NUMBER OF SEQUENCES: 49
CORRESPONDENCE ADDRESS:
ADDRESSEE: Howson and Howson
STREET: Spring House Corporate Center, P.O. Box 457
CITY: Spring House
STATE: PA
COUNTRY: USA
ZIP: 19477
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/136,574A
FILING DATE: 19-Aug-1998
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/932,571
FILING DATE: September 19, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E.
REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: 1997US001/CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-540-9200
TELEFAX: 215-540-5818
TELEX: <Unknown>
INFORMATION FOR SEQ. ID NO.: 47:
SEQUENCE CHARACTERISTICS:
LENGTH: 616 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 47:
US-09-136-574A-47

```

Query Match%      6.6%; Score 109.5; DB 4; length 616;
Best Local Similarity 23.7%; Pred. No. 0.068;
Matches    58; Conservative   24; Mismatches    80; Indels   83; Gaps   13;

QY      34 ASKPYNN-----STSGGSHR-----TGGGGLALGSQVIYDSDGVNPKRYQVKQGDPTVSK 83
           |||:|||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db       65 ASNATFNFNYKLSSGYSCADYILEVFSSGC-----QUPEKOTGD 105
           |||:|||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY      84 IARGLN-----WREIGHINNL--NSSYTIYGQMLTLMSGDLKVEREISSGV 131
           |||:|||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      106 IAVRNKKDKMSNTNOADMWSLQSMTNKGNAKITLY-DGLVLWG-----DEPGAV 157
           |||:|||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY     132 NTAHTPSPVAAOSSRPVOQHBAVGKPTPPVVVVVKKPPTPPVVVQQAPVAVPVTEAP-- 189
           |||:|||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db     158 TPTSTPREVSSAST-----PPRTAT---PPTRESITITPAFTPTP-TTPSPV 199
           |||:|||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY     190 -----FATGSSGVNQFRYPVVGATNPVVRFGELATAVASIYSNMWMFSGRDGDIIN 240
           |||:|||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db     200 TDDTTDDMLFAQGNKIIVDKGKPWLIG--VNMFQFNT---GTNVFDGVMSCNIKSALAE 254
           |||:|||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY     241 ASNAG 245
           |||
Db     255 IANRG 259

RESULT 16
US-08-783-774-2
; Sequence 2, Application US/08783774
; Patent No. 6054130
; GENERAL INFORMATION:
; APPLICANT: Spaete, Richard
; APPLICANT: Jackman, Winthrop
;
```

```

1  TITLE OF INVENTION:  NON-SPLICING VARIANTS OF
2  TITLE OF INVENTION:  GP350/220
3  NUMBER OF SEQUENCES:  19
4  CORRESPONDENCE ADDRESS:
5  ADDRESSEE:  Pennie & Edmonds
6  STREET:  1155 Avenue of the Americas
7  CITY:  New York
8  STATE:  NY
9  COUNTRY:  USA
10 ZIP:  10036/2711
11 COMPUTER READABLE FORM:
12 MEDIUM TYPE:  diskette
13 COMPUTER:  IBM Compatible
14 OPERATING SYSTEM:  DOS
15 SOFTWARE:  FastSEO Version 2.0
16 CURRENT APPLICATION DATA:
17 APPLICATION NUMBER:  US/08/783,774
18 FILING DATE:  15-JAN-1997
19 CLASSIFICATION:  435
20 ATTORNEY/AGENT INFORMATION:
21 NAME:  Cornuzzi, Laura A.
22 REGISTRATION NUMBER:  30,742
23 REFERENCE/DOCKET NUMBER:  7682-037
24 TELECOMMUNICATION INFORMATION:
25 TELEPHONE:  212-798-9090
26 TELEFAX:  212-869-9864
27 TELEX:  6614 PENNIE
28 INFORMATION FOR SEQ ID NO:  2:
29 SEQUENCE CHARACTERISTICS:
30 LENGTH:  907 amino acids
31 TYPE:  amino acid
32 STRANDEDNESS:
33 TOPOLOGY:  unknown
34 MOLECULE TYPE:  protein
35 US-08-783-774-2

```

Query Match 6.5%; Score 109; DB 3; Length 907;
Best Local Similarity 23.2%; Pred. No. 0.13;
Matches 66; Conservative 34; Mismatches 122; Indels 62; Gaps 13;

```

35 SKPTVSTGSGSHRSTSGGGLAIGSYITTSQGVPNY--QVQSGPTSKIAQRYGLNW 92
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
430 TSPITNTTGFAPDPTNTTG-----LPSTHVFNLTLPASTGPTVS-----469

93 REIGHNNINNSYTTIYTQWLT---LWSGDLKXREHSISGVTATHPSPVAVQSSRP 148
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
470 --TADVTPTPAGTTSGASPVTPSPSPMDNOTESKAPDMTSSISPVTTFPNNATSP-- 524

149 VQOHPVQKEP-----PVVVKKPP--TPPVQOAPVAPVPYLEAPATG-----SSGV 137
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
525 ----PAVTTFPNNATSPTPAVTTFPNNATSTLTGKTSPTSAVTTFPNNATSPILGKTSPT 580

198 MGRPVV--GATPVVRRRGCTAVGSGVTSNGMFS--GRGDDLINAGTIVLOADHMD 235
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
581 SAVTTPTPNATSPITLCK--TSPTSAVTTTPTPNATGPTGERSPPQANIN-----HTLG 631

256 GASIVIGHNPGVSSYHIHKDAQVKTGPTVATGRIASMKQPS 239
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
632 GIS-----PTPVVISQPKNATSAVTTGQHNITSSSTSMILRPS 670

Db

```

RESULT 17
US-09-328-599A-1
Sequence 1, Application US/09328599A
Patent No. 6432679
GENERAL INFORMATION:
APPLICANT: MOND, James J. and Lees, Andrew
TITLE OF INVENTION: Enhancement of B Cell Activation by
TITLE OF INVENTION: Co-Ligation of Receptors for Antigen
TITLE OF INVENTION: Using EBV gp350/220 or EBV gp350/220
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &

Matches 63; Conservative 51; Mismatches 149; Indels 47; Gaps 11;

QY 22 FGVTTCILAGCASKRTYNSTSGSGHRTSGGGL-----AIGSOYITDSQVPRVQYK 76
Db 1398 YGIVTSTSGGTPSTSGSGSKYSYINSAASKNNVDPAIIAIVIQESGF--NAKAR 1455
QY 77 QGDTVKIAQRYGLNREIGHINNLSY-----TIYGOULTMSGDLKXRENSISG 130
Db 1456 SGVGANGLMQMPATKSLG-VNNMADPYONMGGTKTIAQLGEGGVEGALAATNAG 1514
QY 133 VNTAHTPSPVAVOSSRPPVOQHPAVOKRTPPVVVVKKRTPPPVVOQAPAPVPTAP 190
Db 1515 -----PGNVIKYGGIPEPKETONVYKIMANYSKLSAISST-----ASY 1555
QY 191 ATGSSGVMOGRYPVGNATNPVRRFGTATVAGSTVTSNGMFPSGRDGDLINASNAQTVOA 250
Db 1556 YTNNSA---FR-----VSKTQOQESGLRSHKGTDPAAKAGTAIKSLQSGKVOIA 1604
QY 251 DHNMD-GASIVYQHTNGFVSYTH-IKDAQYKGTPTVTRGRIASMKQPSGAA-LFEFR 307
Db 1605 GYSKTAGMNVVLIKODGTIVAKTMMMLTPSVKAGSVKAGQITGKVGSTGNSGNHLQ 1664
QY 308 ISRNGVYVDP 317
Db 1665 LEONGKTIID 1674

RESULT 20

US-08-619-812-9
; Sequence 9, Application US/08619812
; Patent No. 6100066
; GENERAL INFORMATION:
; APPLICANT: POTTER, ANDREW A.
; APPLICANT: THEISEN, MICHAEL
; APPLICANT: HARLAND, RICHARD J.
; APPLICANT: RIOUX, CLEMENT R.
; TITLE OF INVENTION: VACCINES FOR HAEMOPHILUS SOMNUS
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: REED & ROBINS
; STREET: 635 BRYANT STREET
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/619,812
; FILING DATE: 15-MAR-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/038,719
; FILING DATE: 29-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: ROBINS, ROBERTA L.
; REGISTRATION NUMBER: 33,208
; REFERENCE/DOCKET NUMBER: 9000-0019,20
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 617-8999
; TELEFAX: (415) 327-3231
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 265 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-619-812-9

Query Match 6.5%; Score 108; DB 3; Length 265;

Best Local Similarity 22.0%; Pred. No. 0.028;

Matches 39; Conservative 35; Mismatches 69; Indels 34; Gaps 5;

QY 73 YQYKQDVTYSKIAQRYGLNREIGHINNLSYTIYGOULTMSG--DIKVERRSISS 129
Db 106 YKVRKGDITMFLIAYISGMDIKELATLNNMSEPHYLSIGQVLIKANNIPDSNMIFPTQITNE 165
QY 130 GVNTHATPSPVAVOSSRPPVOQHPAVOKRTPPVVVVKKRTPPPVVOQAPAPVPTAP 189
Db 166 SEVTONTVNE-TWNANKPTNEQMKPVATPTSTWPKINK---TPPATSNIAIWIW----- 215
QY 190 FATGSSGVMOGRYPVGNATNPVRRFGTATVAGSTVTSNGMFPSGRDGDLINASNAQT 246
Db 216 -----TNGKITIQGFSADG-----NKGIDISGSRGQAVNAAAWT 251

RESULT 21

US-08-619-812-4
; Sequence 4, Application US/08619812
; Patent No. 6100066
; GENERAL INFORMATION:
; APPLICANT: POTTER, ANDREW A.
; APPLICANT: THEISEN, MICHAEL
; APPLICANT: HARLAND, RICHARD J.
; APPLICANT: RIOUX, CLEMENT R.
; TITLE OF INVENTION: VACCINES FOR HAEMOPHILUS SOMNUS
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: REED & ROBINS
; STREET: 635 BRYANT STREET
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/619,812
; FILING DATE: 15-MAR-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/038,719
; FILING DATE: 29-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: ROBINS, ROBERTA L.
; REGISTRATION NUMBER: 33,208
; REFERENCE/DOCKET NUMBER: 9000-0019,20
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 617-8999
; TELEFAX: (415) 327-3231
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 279 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-619-812-4

Query Match 6.5%; Score 108; DB 3; Length 279;

Best Local Similarity 22.0%; Pred. No. 0.03; Mismatches 69; Indels 34; Gaps 5;

QY 73 YQYKQDVTYSKIAQRYGLNREIGHINNLSYTIYGOULTMSG--DIKVERRSISS 129
Db 120 YKVRKGDITMFLIAYISGMDIKELATLNNMSEPHYLSIGQVLIKANNIPDSNMIFPTQITNE 179
QY 130 GVNTHATPSPVAVOSSRPPVOQHPAVOKRTPPVVVVKKRTPPPVVOQAPAPVPTAP 189
Db 180 SEVTONTVNE-TWNANKPTNEQMKPVATPTSTWPKINK---TPPATSNIAIWIW----- 229

QY 190 FATGSSGWMQFRYPVATNPVRRFGTATYAGSTVTSNGMWFSGRDDLINASNAQT 246
Db 230 -----TNGKTIQGFSSADG3-----NKGIDISSRGQAVVAAAAMT 265

RESULT 22

US-08-737-716-13
; Sequence 13, Application US/08737716
; Patent No. 5955258
; GENERAL INFORMATION:
; APPLICANT: Gilbe BUIST
; APPLICANT: Gerard VENEMA
; APPLICANT: Jan KOK
; APPLICANT: Adrianus Marinus LEDBOER
; TITLE OF INVENTION: Process for the lysis of a culture of lactic
; TITLE OF INVENTION: acid bacteria by means of a lysin, and uses of the resulting
; TITLE OF INVENTION: lysed culture.
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pillsbury Madison & Sutro, L.L.P.
; STREET: 1100 New York Avenue, N.W.
; City: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3918
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Microsoft Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/737, 716
; FILING DATE: 22-APR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/NL95/00170
; FILING DATE: 12-MAY-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 94201353.3
; FILING DATE: 12-MAY-1994
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 671 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus faecalis
; IMMEDIATE SOURCE:
; CLONE: Fig.5a (S. faecalis)
; US-08-737-716-13

Query Match 6.4%; Score 107.5; DB 2; Length 671;
Best Local Similarity 20.8%; Pred. No. 0.12;
Matches 66; Conservative 43; Mismatches 77; Indels 131; Gaps 15;

QY 34 ASKPTVNS-----TSGSGSHRTSGSGGLALGQVITTSQGVPRNRYQKOG 78
Db 311 ATDPSTYAKLNNTAVNTALTYDTPSSGNN-TGGTVNPETGGSNNQSGTNYTYVXSG 368
QY 79 DTASKIAQRYGL-----NMREI-----GHINNLNSY 105
Db 369 DTINKIAAQYGVSVANLRSMNGISGDLIFVQKLIYKKGASGNTGSGNGSGNNQSGTN 428
QY 106 TTYTGQMLTMSGDLKVRERSISSGVNTAHTPSFVAVQSSRPVQGHAVQKPTPVVVV 165
Db 429 TTY-----TVKSGD-TINKIAAQYGVTVANLRSMNGISGDLIFVQK-----LIV 472
QY 166 KKPTPTPVVQGPAPVAPVTEAPFATGSSGWMQFRYPVATNPVRRFGTATYAGSTVT 225
Db 473 KKG-----SGNTG-----GSSNN 491
QY 226 SNG--MWFSGRDDLINASNAQTIVIQADHMDGASIVIQHTNGFVSSYIHKDAQVKTGD 283

Db 492 QSGTNTYTYTKSGDITLTK-----IAAQYGVSVAN-LRSMNGI-----SGD 530
QY 284 TVRTGORIAMKNQPSG 300
Db 531 LIFAGQKII-VKKGTSG 546

RESULT 23

US-09-109-841-2
; Sequence 2, Application US/09109841
; Patent No. 6207436
; GENERAL INFORMATION:
; APPLICANT: Bjornvad, Made E.
; APPLICANT: Hatakeyama, Mariko
; APPLICANT: Schulten, Martin
; APPLICANT: Nielsen, Jack B.
; TITLE OF INVENTION: Endo-Beta-1, 4-Glucanases From
; TITLE OF INVENTION: Saccharothrix
; FILE REFERENCE: 5195.200-US
; CURRENT APPLICATION NUMBER: US/09/109, 841
; EARLIER FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: 0812/97
; EARLIER FILING DATE: 1997-07-04
; EARLIER APPLICATION NUMBER: 0846/97
; EARLIER FILING DATE: 1997-07-11
; EARLIER APPLICATION NUMBER: 60/053, 506
; EARLIER FILING DATE: 1997-07-23
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 490
; TYPE: PRN
; ORGANISM: Saccharothrix australiensis
; US-09-109-841-2

Query Match 6.4%; Score 107; DB 4; Length 490;
Best Local Similarity 20.4%; Pred. No. 0.084;
Matches 77; Conservative 41; Mismatches 137; Indels 122; Gaps 16;

QY 7 INSQOKPI-KRIGLFGVITTCILAGCASKPTYNSTSGSHRTS-----GSGGLA 57
Db 1 MHRPSKRLPTRKRVPAVAVAGTVLAGGVALTSTNIAQAAGCRVDYAVTSQMPGFGAA 60
QY 58 IG-----SQVITDSQGVPRNRYQ-----VQ-----GPTVSKIAQRYG 89
Db 61 VVTNLGDLPISSWELSWTPPDGGVQOQLMNGVHSTSGSNVTYKEMSGSVGTNASVQVG 120
QY 90 LNMREIGHINNLNSSYTIYQGM/LTMSGDLKVRERSISSGVNTAHTPSFVAVQSSRPV 149
Db 121 FN-----GSMNGANNAPTSFTLNG-TSCNGAV-----CGPTTEPTPEPTPEPTPEPTP 167
QY 150 QQHP-AVQKPTPVVVKKPTPTPVVQGPAP-VAPVTEAPFATGSSGWMQFRYPVAT 207
Db 168 EPTPEPTPEPTPEPTPEPTPEPTPEPTPEPTPEPTPEPTPEPTPEPTPEPTPEPTPEPT 220
QY 208 NPNVRRFGTATYAGSTVTSNGMWFSGRDDL-----NASAGTIVIQADHMD----- 255
Db 221 NOSRYAMQAA-----SSGDKDLAKIALITPOAYVWGMNNEASHAQOEVRDI 266
QY 256 -----GASIVI-----OHTNGFVSSYIHK-----DAQ 278
Db 267 TSAAMAAAGRTAVLVVVAIPGRDGGHSSGSGVSTSEVAQWIDTYAQAQIVGNPVLDPDAL 326
QY 279 VKTGDVTRTGQRIASMK 295
Db 327 PMLGDCDGGQDRVGFLLK 343
RESULT 24
US-09-556-706B-2
; Sequence 2, Application US/09556706B
; Patent No. 6458364

GENERAL INFORMATION:
APPLICANT: Spaete, Richard
TITLE OF INVENTION: JACKMAN, WINCHROP
FILE REFERENCE: 7682-050-999
CURRENT APPLICATION NUMBER: US/09/556,706B
CURRENT FILING DATE: 2000-04-24
PRIOR FILING DATE: 1997-01-15
PRIOR APPLICATION NUMBER: 08/229,291
PRIOR FILING DATE: 1994-04-18
NUMBER OF SEQ ID NOS: 19
SOFTWARE: Patent in version 3.0
SEQ ID NO 2
LENGTH: 878
TYPE: PRT
ORGANISM: Virus
FEATURE:
OTHER INFORMATION: gp350
US-09-556-706B-2

Query Match 6.4%; Score 106.5; DB 4; Length 878;
Best Local Similarity 22.1%; Pred. No. 0.22;
Matches 63; Conservative 29; Mismatches 118; Indels 75; Gaps 12;

35 SKPTNNTSGSGSHRTSGSGGLAISQVITDSQGVNRY--QVKGDTVSKIAQRYGLNW 92
430 TSPITLNTGFAADPNTTGT-----LPSSTHVPITLTPASTGPTVS----- 469
93 REIGHINNLSSYTYTGQMLT---LMSGDLKVERSSISSGVNTAHTPSPFAVQSSRP 148
470 --TADVTPTPTGTTSGASPVTPSPMDNGTESKAPDMTSTSPVTPPTNATSPPT--- 524
149 VOQHAPVOKPTP---PVVVVKKPTP--TPVVQOPAPVAPVTEAPFATGSSGVQFRY 202
525 ---FAVTPPTNATSPPTAVTTPPTNATSPPTLGTSTSAVTTPTNATSPPT----- 573
203 PVGATNPVRRFGTAVAGSTVTSNGMWFSGRDGDLINASVAGVVIQADHNMGASL--- 259
574 -LGKTSPT-----TSAVTTPTPTP-----ATSPFLGKTSPTSAVTTPTNATSPPTVGT 619
260 ---VIQHTNGFVSSYHIIHDAQVKTGDTVTCGORLASMKNOPS 299
620 SFOANATNHTLGTSTPTVVTSPKRNATSAVTTGQH-----NRPS 659

RESULT 25
US-08-155-171B-15
Sequence 15, Application US/08155171B
Patent No. 5543264
GENERAL INFORMATION:
APPLICANT: Anderson, Carl W.
TITLE OF INVENTION: Co-Factor Activated Recombinant
TITLE OF INVENTION: Adenovirus Proteins (As Amended)
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Millitia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: USA
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/155,171B
FILING DATE: 19-NOV-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/851,217
FILING DATE: 13-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/545,585
FILING DATE: 29-JUN-1990
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: BNL91-01A2, AUI93-22
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 861-9540
TELEFAX: (617) 861-9540
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 250 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-155-171B-15

Query Match 6.2%; Score 103.5; DB 1; Length 250;
Best Local Similarity 24.0%; Pred. No. 0.067;
Matches 42; Conservative 27; Mismatches 39; Indels 67; Gaps 9;

82 SKIAQRYGL-----NMRREIGHINNLSSYTYTGQMLTMSGDL----- 120
7 ASLAPRHGRSPFMGNWDIGTSMNSGAFS-----KSLMSGIKRNGSTIKNTGSAKNS 61
121 -----KVERSS-----ISSGVNTAHTPSPFAVQSS-----RPVQO-HPAVQ 156
62 STGQMLRKLKEQNFQKQVNDGLASGISGVVDLANQAVQKINSKIDPRPVEPPVAE 121
157 KETPPVVVKKPTP-----TPP-----VQOPAPVAPVTEAPFATGSSG 196
122 TVSPBGRKPKRPDRRETLVTQIDEPFSYEAKQGLPTTRPI--APMATGVLG 174

RESULT 26
US-08-435-998-15
Sequence 15, Application US/08435998
Patent No. 5935840
GENERAL INFORMATION:
APPLICANT: Anderson, Carl W.
TITLE OF INVENTION: Co-Factor Activated Recombinant
TITLE OF INVENTION: Adenovirus Proteins (As Amended)
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Millitia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: USA
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/435,998
FILING DATE: 05-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/155,171
FILING DATE: 19-NOV-1993
APPLICATION NUMBER: US 07/851,217
FILING DATE: 13-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/545,585
FILING DATE: 29-JUN-1990
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia

REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: ENL91-01A2, AUI93-22
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 861-6240
TELEFAX: (617) 861-9540
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 250 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-435-998-15

Query Match 6.2%; Score 103.5; DB 2; Length 250;
Best Local Similarity 24.0%; Pred. No. 0.067;
Matches 42; Conservative 27; Mismatches 39; Indels 67; Gaps 9;

QY 82 SKIAQRVGL-----NMREIGHINNLSSTYITYGQMLTMSGDL-----120
DB 7 ASIAPRHGRPRPMQNDIGTSNMSGAFS-----MGLMSGIKNFGSTIKYGSKAMNS 61
QY 121 -----KREBS-----ISSGVNTAHTSPVAVQSS-----RPPVQO-HPAVQ 156
DB 62 STQQLMDKDKKEQNFQOKVVDGLASGISGVVDLANQVOKINSKLDPRPVEPEPPAVE 121
QY 157 KRPPIVVVKKPFP-----RPP-----VVQGPAPVAPVTEAPFATGSSG 196
DB 122 TVSPBGRGRPRPRDRETLVTQIDPEPPEALKQGLPTTRPI--APMATGVLG 174

RESULT 27

US-08-217-327-4
Sequence 4, Application US/08217327
Patent No. 5474925
GENERAL INFORMATION:
APPLICANT: John, Maliyakal E
APPLICANT: Barton, Kenneth A
TITLE OF INVENTION: Immobilized Proteins in Cotton Fiber
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Quarles and Brady
STREET: P.O. Box 2113
CITY: Madison
STATE: WI
COUNTRY: USA
ZIP: 53701-2113
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/217,327
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/812,233
FILING DATE: 19-DEC-1991
ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J
REGISTRATION NUMBER: 27,386
REFERENCE/DOCKET NUMBER: 1122990831
TELECOMMUNICATION INFORMATION:
TELEPHONE: 608-251-5000
TELEFAX: 608-251-9166
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 214 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-217-327-4

Query Match 6.2%; Score 103; DB 1; Length 214;
Best Local Similarity 42.9%; Pred. No. 0.06;
Matches 27; Conservative 6; Mismatches 26; Indels 4; Gaps 3;

QY 132 NTAHTSPVAVQSSRPVQOHPAVQKTPPVVVKKPTP--PPVQOPAPVA--PPTAP 189
DB 32 STPATPTPPA--STPPTTQAPPTPTATPPPVSTPPTSPPPVTASPPVSTPSSSP 89
QY 190 FAT 192
DB 90 PAT 92

RESULT 28

US-09-120-927-2
Sequence 2, Application US/09120927
Patent No. 6262018
GENERAL INFORMATION:
APPLICANT: Kim, Jihyun Francis
APPLICANT: Beer, Steven V.
TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR FROM
TITLE OF INVENTION: ERWINIA AMILOVORA AND ITS USE
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
STREET: P.O. Box 1051, Clinton Square
CITY: Rochester
STATE: New York
COUNTRY: U.S.A.
ZIP: 14603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/120,927
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/055,108
FILING DATE: 06-AUG-1977
ATTORNEY/AGENT INFORMATION:
NAME: Goldman, Michael L.
REGISTRATION NUMBER: 30,727
REFERENCE/DOCKET NUMBER: 19603/1581
TELECOMMUNICATION INFORMATION:
TELEPHONE: (716) 263-1304
TELEFAX: (716) 263-1600
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 447 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-120-927-2

Query Match 6.1%; Score 102; DB 4; Length 447;
Best Local Similarity 23.1%; Pred. No. 0.21;
Matches 77; Conservative 51; Mismatches 147; Indels 59; Gaps 16;

QY 6 AINSQOKPKRGL-IFGVITTCILAGCAKPTVNSTSGSGH-RTSG---SGGL-----56
DB 30 ANSALGQOPDRQTEHQMAQLAEILKLSLSPSGNAATGAGGNDQTTGVGNAGLNGRK 89
QY 57 -AIGSVITDSQGVPRVQVQGVTVSKIAQRYGLNMRREIGHINNLSSTYI-----Y 108
DB 90 GTAGTTPGDSQ---NMLSEMGNGGLDQAITPDGGGGGIGNPLKAMLIKLIARMDQ 146
QY 109 TQQMLTMSGDLKVRERSISSGVNTAHTSPVAVQSSRPVQOHPAVQKTPPVVVKKP 168
DB 147 SDQFGQPGTGNNSASGTSSSGSPNDLSGKAPSGNSPSGNSVSTFSPP-----S 200

QY 169 TPTTPVVOQAPAPV-----PYTEAPFATSSGVMOFRYVGATNPVRRFGTA--- 217
DB 201 TPTSPSPDPPSPKAGSGSTPVTDHDPVGSAG-----IGAGNSVA--FTSAGAN 251
QY 218 -TVAGSTVT-SNGMFGSRDGLINASNAGTIVIQADHNM-----DGASI--VIOHTNGF 267
DB 252 QTVLHDTITVKAGGVFDGKQFTAGSELGDSGSENGKPLFIEDGASLKNVTMGDDG- 310
QY 268 VSSYIHI-KDAQVKTGDTVRTGQRIASMKMCPSG 300
DB 311 -ADGILHYGDAKIDMLHVTNVGEDAITYKPNAG 343

RESULT 29

US-08-276-213-3
Sequence 3, Application US/08276213
Patent No. 5536655
GENERAL INFORMATION:
APPLICANT: Thomas, Steven
APPLICANT: Laymon, Robert
APPLICANT: Himmel, Michael
TITLE OF INVENTION: GENE ENCODING FOR THE E1 ENDOGLUCANASE
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSER: National Renewable Energy Laboratory
STREET: 1617 Cole Boulevard
CITY: Golden
STATE: CO
COUNTRY: USA
ZIP: 80401-3993
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/276, 213
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: O'Connor, Edna
REGISTRATION NUMBER: 29, 252
REFERENCE/DOCKET NUMBER: NREL IR# 94-08
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303)231-1000
TELEFAX: (303)231-1098
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 521 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYDROTHERMAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
US-08-276-213-3

Query Match 6.1%; Score 101.5; DB 1; Length 521;
Best Local Similarity 25.2%; Pred. No. 0.3;
Matches 53; Conservative 20; Mismatches 71; Indels 66; Gaps 12;
QY 111 QWLTLMSS-----GDLKYRERSISSGVNTAHTP-----SPVAVQSSRPVVOQHAPV 155
DB 315 QW-TFWMNDPSDGTGLKDMQVTVVDGYLAPKSIKFPVGSAS-FSSQSPSEV 372
QY 156 Q-KTTPVVVVKETPTTPVVOQAPV-----APPVTEAPATGSSGV----- 198
DB 373 SPSPSPSPASARPTPTPTASPTPLPTATPTPTASPTSPASGARCTASQVNS 432
QY 199 -----QEFYYPVGAATNPVRRFGTAVAGSTVTSNGMFGSRDGLINASNAGTIVIQADHNM 254

DB 433 DMGNFPTVAATN-----SGSVATKTIWVS--WTEGNGQITLNSMNAV-----TQ 477
QY 255 DGASI-----YIO-----HTNGFVSSY 271
DB 478 NGQSVTARMMSYNNVYIQPGQNTTFGFGASY 507

RESULT 30

US-08-728-323A-2
Sequence 2, Application US/08728323A
Patent No. 5948676
GENERAL INFORMATION:
APPLICANT: Chang, Yuan
APPLICANT: Bohenzky, Roy A.
APPLICANT: Russo, James J.
APPLICANT: Edelman, Isidore S.
APPLICANT: Moore, Patrick S.
TITLE OF INVENTION: Immediate Early Protein From Kaposi's
TITLE OF INVENTION: Sarcoma-Associated Herpesvirus, DNA
ENCODING SAME AND USES THEREOF
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSER: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/728, 323A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28, 678
REFERENCE/DOCKET NUMBER: 0575/52268/JPM/MSK/SKS
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-278-0400
TELEFAX: 212-391-0525
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1162 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-728-323A-2

Query Match 6.1%; Score 101.5; DB 2; Length 1162;
Best Local Similarity 24.0%; Pred. No. 0.95;
Matches 40; Conservative 12; Mismatches 42; Indels 73; Gaps 8;
QY 127 ISSGVNTAHTPSPVAVQSSRPVVOQHAPVQKPT-----PPVYVVKKP-----TPT 171
DB 120 VSPPTDTHSPSPALPPTGSPSSQRPPLSPTRGRDSTPMRPPSCOTTFPHSTPTP 179
QY 172 P-----PVVOQAPVAPV----- 184
DB 180 PEPSPKSSPDLAPSTLRSLKRLSSPOGPTLNPICSP-FVSPFRCDFPANSVYPPW 238
QY 185 VTAAPFATGSSGVMOFRYVGATNPVRRFGTAVYA-GSTVTSNGMW 230
DB 239 AHSFPIVSSS-----DGDTPP--RQPTSPISISSPSSEGSW 276
RESULT 31
US-09-298-568-2
Sequence 2, Application US/09298568
Patent No. 6322792

REFERENCE/DOCKET NUMBER: 6523-003

SEQUENCE CHARACTERISTICS:
LENGTH: 268 amino acids

SEQUENCE CHARACTERISTICS:
LENGTH: 268 amino acids

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-431-387-4

Query Match
Best Local Similarity 24.9%; Score 98.5; DB 1; Length 268;
Pred. No. 0.22;
Matches 71; Conservative 46; Mismatches 95; Indels 73; Gaps 18;

QY 2 TTTAINSONOKFKIKLIGF-----VITTCI-----LAGCA-----SKPTNYSNG 44
DB 2 TVVWGISFINTQOAHNRG-IFGNGARVALDITGIAHTPDLRIAGASISSEPSYHDNG 60
QY 45 SCSHRTSGGGLAIGSQVITDSQV-----FNRQVQGD-----TVSKIAQRYGLNMR 93
DB 61 HGTI---VAGTIA---ALNNSIGVLGVAPSADLVAVYLDNRSGSGLASVAG-GLIEMA 111
QY 94 EIGHINNLSSTYTYGQWLTLSGDLKVR-ERISISGV-----NT-----AHTP 137
DB 112 INNMMHIIIMSLGISTG-----SSTELAVNRKANNAGILLVGAAGNTGROGVNYPARYS 165
QY 138 SPVAV---OSSRPP---VOGHAVOKEPTPPVYVVKKPTPPVVOQAPVAPVTEAPFA 191
DB 166 GWAVAAVDQNGQPSFSTYGEITISAPGVANSTYGNRYVSLGISTMA-----TPHV 220
QY 192 TSSGVMORRYVAGTNPVVRFFG-TATVAGS-TVTSNGMWFSGR 234
DB 221 AGVAAVKSRYPSYTNNOIKRINQIATATYLSPLGYGGLVHAGR 265

RESULT 34

US-09-165-239A-4
Sequence 4; Application US/09165239A
Patent No. 6344554
GENERAL INFORMATION:
APPLICANT: JOHNSON, ALEXANDER
APPLICANT: BRAUN, BURKHARD R
TITLE OF INVENTION: POLYNUCLEOTIDE SEQUENCES FROM CANDIDA
TITLE OF INVENTION: ALBICANS ENCODING POLYPEPTIDES ASSOCIATED WITH FILAMENTOUS
FILE REFERENCE: 220022000700
CURRENT APPLICATION NUMBER: US/09/165,239A
PRIOR FILING DATE: 1998-10-01
PRIOR APPLICATION NUMBER: 60/068,065
NUMBER OF SEQ ID NOS: 13
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO 4
LENGTH: 750
TYPE: PRT
ORGANISM: Candida albicans
US-09-165-239A-4

Query Match
Best Local Similarity 31.5%; Score 98.5; DB 4; Length 750;
Pred. No. 0.96;
Matches 46; Conservative 12; Mismatches 65; Indels 23; Gaps 7;

QY 128 SSGNTAHTSPVAVOGSSRPVVOQHAPVQKPTPPVYVVKKPTPPVVOQAPVAP--PV 185
DB 603 SSPVAPGTSSSPVAPSSAPATESAPATE--SSPVAPGTETTPATPGAES--TPVAPVAP 659
QY 186 TEAPFATGSSGVQMFYRYPVATNPVVRFGTATVAGSTVTSNMWPSGSDGLINASNG 245
DB 660 SSAF-AVESSPVA---PGVETTPV-----APVASTAKTSALVSTTEGILPTLSV 708
QY 246 TVIQADHNMDCASIVIOHTNGFVSSY 271
DB 709 PAIOPSANS-----SYTIASVSSF 727

RESULT 35
US-08-296-791-5
Sequence 5; Application US/08296791

Patent No. 6245337
GENERAL INFORMATION:
APPLICANT: St Geme III, Joseph W.
APPLICANT: Raikov, Stanley
TITLE OF INVENTION: Haemophilus Adherence and Penetration
TITLE OF INVENTION: Protein
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSER: Flehr, Hobbach, Test, Albritton & Herbert
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/296,791

FILING DATE: 25-AUG-1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Treccartin, Richard F.

REGISTRATION NUMBER: 31,801

REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 781-1989

TELEFAX: (415) 398-3249

TELEX: 910 277299

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 1702 amino acids

TYPE: amino acid

TOPOLOGY: unknown

US-08-296-791-5

Query Match
Best Local Similarity 23.0%; Score 98.5; DB 4; Length 1702;
Pred. No. 3.1;
Matches 42; Conservative 24; Mismatches 62; Indels 55; Gaps 8;

QY 38 TYNSTSGSGH-----RTSGSGGLAIGSQVITDSQGVNRYQVQGTIV 81
DB 896 TVNLSGNGSFYVLTDLNKGDKVVTKSGTGNFTL--GVADKGTGPTKMTLTFDAS 952
QY 82 SKIAQRYGLNMRIGH-----INNLSSTYTYGQWLTLSGDLKVERISISGV 131
DB 953 N--ATRNINLAVSLVGTVDIGAWKYLRYVNRGRYDLNPE-----VEKR--NQTIV 998
QY 132 NTAHTPSPVAVOGSSRPVVOQHAPVQKPTPPVYVVKKPTPPVVOQAPVAPVTEAPFA 191
DB 999 DTNITTPNNIQADVDSV-----PSNMEDIAVEITPVP--PAPATPSETTETIVA 1046
QY 192 TGS 194
DB 1047 ENS 1049

RESULT 36

PCT-US95-10661A-5
Sequence 5; Application PCT/US9510661A
GENERAL INFORMATION:
APPLICANT: Washington University, et al.
TITLE OF INVENTION: Haemophilus Adherence and Penetration Protein
TITLE OF INVENTION: 9
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSER: Flehr, Hobbach, Test, Albritton & Herbert
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10661A
FILING DATE: 16-AUG-1995
CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/296,791
FILING DATE: 25-AUG-1994
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Treacatin, Richard F.
REGISTRATION NUMBER: 31,801
REFERENCE/DOCKET NUMBER: RP-59941/RPT
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1702 amino acids
TYPE: amino acid
TOPOLOGY: unknown
PCT-US95-10661A-5

Query Match 5.9%; Score 98.5; DB 5; Length 1702;
Best Local Similarity 23.0%; Pred. No. 3.1;
Matches 42; Conservative 24; Mismatches 62; Indels 55; Gaps 8;

QY 38 TYNSTSGSGSH-----RTSGSGGLAIGSVITPDSGVNRYOVKQDITV 81
DB 896 TYNLSGNSFFYITLDSKNGDKVVTYSATNGFTL--QVADKTEPTKNEITLFDAS 952
QY 82 SKIAQRVGLMREIGH-----INNLSYTYITGOMLTLMSGDLKVRERSISSGV 131
DB 953 N--ATNNNLVSLVGMTVDLGAMKYLKRVNNGRYDLNPE-----VEKR--NQTV 998
QY 132 NTNHTSPVAVGSSRPVQCHPAVQKTPPVVVVKKTPPPPVVQOPAPVAPVTEAPFA 191
DB 999 DTNNITTPNNIQADVDSV-----FSNNEIARVETTPVP-----PAPATPSETTETVA 1046
QY 192 TGS 194
DB 1047 ENS 1049

RESULT 37
US-08-478-507-7
Sequence 7, Application US/08478507
Patent No. 6120988
GENERAL INFORMATION:
APPLICANT: Reyes, Gregory R
APPLICANT: Yarbough, Patrice O
APPLICANT: Bradley, Daniel W
APPLICANT: Krawczynski, Krzysztof Z
APPLICANT: Tam, Albert
APPLICANT: Fry, Kirk B
TITLE OF INVENTION: DNA Sequences of Enterically Transmitted
TITLE OF INVENTION: No. 6120988-A/No. 6120988-B Hepatitis Viral Agent
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dehlinger & Associates
STREET: 350 Cambridge Avenue, Suite 250
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/478,507
FILING DATE: 07-JUN-1995
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/279,823
FILING DATE: 25-JUL-1994
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/681,078
FILING DATE: 05-APR-1991
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/505,888
FILING DATE: 05-APR-1990
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/420,921
FILING DATE: 13-OCT-1989
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/367,486
FILING DATE: 16-JUN-1989
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/336,672
FILING DATE: 11-APR-1989
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/208,997
FILING DATE: 17-JUN-1988
ATTORNEY/AGENT INFORMATION:
NAME: Sholtz, Charles K.
REGISTRATION NUMBER: 38,615
REFERENCE/DOCKET NUMBER: 4600-0183.22
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 324-0880
TELEFAX: (650) 324-0960
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1693 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-478-507-7

Query Match 5.9%; Score 98; DB 3; Length 1693;
Best Local Similarity 24.9%; Pred. No. 3.4;
Matches 43; Conservative 14; Mismatches 62; Indels 54; Gaps 9;

QY 96 GHI-----NNLSYTYITGOMLTLMSGDLKVRERSISSGVNTNHTPS-PVAVGSSRPVQ 150
DB 685 GHVESANPFCGSESTLYRTW-----SEVDVSSPARPDIGFMSSEPSIP 728
QY 151 QH---PVOKPPPVVVVKKPPTPPVQOPA-----PVA PVTAEAPFATGSSGVMQFRYP 203
DB 729 SRAATPTLAPLDPDPDPSPPSPALAPASGATGAPATH---QTRRRRLFTYP 785
QY 204 VGATNPNVRRFGTATVAGSTVTSNGMWFSGRQDGLINASNAGTVIOADHMDG 256
DB 786 -----DGSKVFAAGSLFESTCTW-----LVNASN-----VDHRRPGG 815

RESULT 38
US-09-128-275A-7
Sequence 7, Application US/09128275A
Patent No. 6229005
GENERAL INFORMATION:
APPLICANT: Reyes, Gregory R
APPLICANT: Yarbough, Patrice O
APPLICANT: Bradley, Daniel W
APPLICANT: Krawczynski, Krzysztof Z
APPLICANT: Tam, Albert
APPLICANT: Fry, Kirk B
TITLE OF INVENTION: DNA Sequences of Enterically Transmitted
TITLE OF INVENTION: No. 6229005-A/No. 6229005-B Hepatitis Viral Agent
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:

ADDRESSER: Dehlinger & Associates
STREET: 350 Cambridge Avenue, Suite 250
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/128,275A
FILING DATE: 03-AUG-1998
APPLICATION DATA:
APPLICATION NUMBER: US 08/279,823
FILING DATE: 25-JUL-1994
APPLICATION DATA:
APPLICATION NUMBER: US 07/681,078
FILING DATE: 05-APR-1991
APPLICATION DATA:
APPLICATION NUMBER: US 07/505,888
FILING DATE: 05-APR-1990
APPLICATION DATA:
APPLICATION NUMBER: US 07/420,921
FILING DATE: 13-OCT-1989
APPLICATION DATA:
APPLICATION NUMBER: US 07/367,486
FILING DATE: 16-JUN-1989
APPLICATION DATA:
APPLICATION NUMBER: US 07/336,672
FILING DATE: 11-APR-1989
APPLICATION DATA:
APPLICATION NUMBER: US 07/208,997
FILING DATE: 17-JUN-1988
ATTORNEY/AGENT INFORMATION:
NAME: Petithory, Joanne R.
REGISTRATION NUMBER: 42,995
REFERENCE/DOCKET NUMBER: 4600-0183.24
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 324-0960
TELEFAX: (650) 324-0980
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1693 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-128-275A-7

Query Match 5.9%; Score 98; DB 4; Length 1693;
Best Local Similarity 24.9%; Pred. No. 3.4;
Matches 43; Conservative 14; Mismatches 62; Indels 54; Gaps 9;

QY 96 GHI-----NNLNSSTYITGQWLTLMGDLKVRERSISSGVNTAHTPS-PVAVOSSRPVQ 150
DB 685 GHVESANPFCGESSTLTTRW-----SEVDAVSSPARPDLGFMSESPISIP 728
QY 151 QH---PAVOKRTPPVVVVKKPTPTPPVVOQA---PVAPVTEAPFATGSSGVMOFRYP 203
DB 729 SRAATPTLAAPLPPAPDPSPPSAPALAEFASGATGAPATH---QTARRHRLIFTYP 785
QY 204 VGATNPVRRFGTAVAGSTVTSNGMFMFSGRDGDLINASVAGVIVIQADHMDG 256
DB 786 -----DGSKVFAGSLFESEICTW-----LVNASN-----VDHRPGG 815

RESULT 39
US-09-553-427-7
Sequence 7, Application US/09553427
Patent No. 6379891
GENERAL INFORMATION:
APPLICANT: Reyes, Gregory R

APPLICANT: Yarbough, Patrice O
APPLICANT: Bradley, Daniel W
APPLICANT: Krawczynski, Krzysztof Z
APPLICANT: Tam, Albert
APPLICANT: Fy, Kirk E
TITLE OF INVENTION: DNA Sequences of Enterically Transmitted
NUMBER OF SEQUENCES: No 6379891-A/No. 6379891-B Hepatitis Viral Agent
CORRESPONDENCE ADDRESS:
ADDRESSER: Dehlinger & Associates
STREET: 350 Cambridge Avenue, Suite 250
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/553,427
FILING DATE:
APPLICATION DATA:
APPLICATION NUMBER: US 08/478,507
FILING DATE: 07-JUN-1995
APPLICATION DATA:
APPLICATION NUMBER: US 08/279,823
FILING DATE: 25-JUL-1994
APPLICATION DATA:
APPLICATION NUMBER: US 07/681,078
FILING DATE: 05-APR-1991
APPLICATION DATA:
APPLICATION NUMBER: US 07/505,888
FILING DATE: 05-APR-1990
APPLICATION DATA:
APPLICATION NUMBER: US 07/420,921
FILING DATE: 13-OCT-1989
APPLICATION DATA:
APPLICATION NUMBER: US 07/367,486
FILING DATE: 16-JUN-1989
APPLICATION DATA:
APPLICATION NUMBER: US 07/336,672
FILING DATE: 11-APR-1989
APPLICATION DATA:
APPLICATION NUMBER: US 07/208,997
FILING DATE: 17-JUN-1988
ATTORNEY/AGENT INFORMATION:
NAME: Sholtz, Charles R.
REGISTRATION NUMBER: 38,615
REFERENCE/DOCKET NUMBER: 4600-0183.22
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 324-0960
TELEFAX: (650) 324-0980
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1693 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-553-427-7

Query Match 5.9%; Score 98; DB 4; Length 1693;
Best Local Similarity 24.9%; Pred. No. 3.4;
Matches 43; Conservative 14; Mismatches 62; Indels 54; Gaps 9;

QY 96 GHI-----NNLNSSTYITGQWLTLMGDLKVRERSISSGVNTAHTPS-PVAVOSSRPVQ 150
DB 685 GHVESANPFCGESSTLTTRW-----SEVDAVSSPARPDLGFMSESPISIP 728
QY 151 QH---PAVOKRTPPVVVVKKPTPTPPVVOQA---PVAPVTEAPFATGSSGVMOFRYP 203
DB 729 SRAATPTLAAPLPPAPDPSPPSAPALAEFASGATGAPATH---QTARRHRLIFTYP 785

QY 204 VGATNPVRRFGTATVAGSTVTSNGMWFSGRDGLINASNAGTVIQADHNMDS 256
Db 786 -----DGSKVFAAGSLFESTCTW-----LVNASN-----VDHRRPGG 815

RESULT 40

US-09-579-181-2
; Sequence 2, Application US/09579181
; Patent No. 6365372
; GENERAL INFORMATION:
; APPLICANT: Chiviva, John
; APPLICANT: Yaciuk, Peter
; TITLE OF INVENTION: SNF2 Related CBP Activator Protein (SRCAP)
; FILE REFERENCE: 16153-4247
; CURRENT APPLICATION NUMBER: US/09/579,181
; CURRENT FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: 60/136,620
; PRIOR FILING DATE: 1999-05-27
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 2972
; TYPE: PRT
; ORGANISM: Human
US-09-579-181-2

Query Match

5.8%; Score 97; DB 4; Length 2972;

Best Local Similarity 25.7%; Pred. No. 9.6; Matches 52; Conservative 20; Mismatches 52; Indels 78; Gaps 12;

QY 137 PSPVAVQS--SRPVOQHPAVQKPTPPVVVVKKPTTPPVVQCPAPVAPVTEAPFATGS 194
Db 890 PSPGLVLSGTSRP-----PTPTLSL--KTPPAPVRLSPAP--PP-----GS 927
QY 195 SGVMQ-FRIYVGATNPVRRFGTATVAGSTVTSNGMWFSGRDGLINASNAGT-----V 247
Db 928 SSLKPLTVPPGGYFPAPAAITTTTATATTT-----AVPAPTPAPQRLI 973
QY 248 IQADHNMDSASIVIQHTNGFVSYIHKDAQVKGTGTVRTGQRIASMKNOP---SGAL 303
Db 974 LSPD-----MQARLPSGEVVSIGQ-LASLAQRVANNAGGSKP 1009
QY 304 FEPRIERN-----GVYDPLTV 320
Db 1010 LTFQIQGNKLLTGAQVQLAV 1031

Search completed: July 8, 2003, 11:08:55
Job time : 30 secs

THIS PAGE BLANK (USPTO)

Db 269 AGDALRYGVNLIILIKHNDSTLAVAHNESILVKKDOOEKAGQOIAKMGSSGTWIKLHFE 328
QY 308 ISRNGVYDPLTVL 321
Db 329 IRYGQGVDPMRYL 342

RESULT 2

US-10-156-761-9179
; Sequence 9179, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: HORIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 9179
; LENGTH: 301
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-9179

Query Match 10.9%; Score 182.5; DB 9; Length 301;
Best Local Similarity 24.3%; Pred. No. 9.8e-07;
Matches 83; Conservative 35; Mismatches 107; Indels 117; Gaps 15;

QY 46 GSHR-----TSGGGGLAI-----GSGVITDSCGVRN----- 72
Db 5 GKHRRRSRPRFTRSIIVAGTGGALPLPMGATGAIAATPIAIVSCKVSAALVAAKQGA 64
QY 73 -----YQVKODPTSKIAQRYGLMWRREIGHINNINSYTTTGTGMLTMSGDLKVRER 125
Db 65 EKSQGTIVAVRABDSLSKIADQSV-----TGGMKKLYSDN-----R 101
QY 126 SISSGVNT-----AHTPSFVAVOSSRPPVOQHAPVOKPTPPVVVVKKPTPTPPV 174
Db 102 SAIGDPTLIHPGLKLTIGAKSASSAATQS-----TATKPAATGVKSAT----- 145
QY 175 VQOPAFVAPVTEAPATSSGVMQFRYPVGAIVNRRFGATVAGSTVTSNGMWFSG- 233
Db 146 APTFASKITTATRAADVTYTAG---YTLFVDDG-----ATIGTAVKTIAGSWSSG 192
QY 234 -----RDGDLINASNAGTVIOAD-HNMDGASIVIOHTNGFVSYYHIKDAQVKTGDT 284
Db 193 HTGVDFVVPVGTITIKAVAGTIVSAGMGAYGNEVVVRHADQYQYAHMSQLSVSTQS 252
QY 285 VRTGOR--TASMKNOPSGAAL-FEPRISHN-GVYVDPPLTVLK 322
Db 253 VAEGRQLGSGATGVNTGPHLHFEIRTPSYGSDVDPAVYLR 294

RESULT 3

US-10-156-761-12234
; Sequence 12234, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: HORIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI

; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 12234
; LENGTH: 203
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-12234

Query Match 8.5%; Score 142; DB 9; Length 203;
Best Local Similarity 37.4%; Pred. No. 0.00092;
Matches 43; Conservative 15; Mismatches 39; Indels 18; Gaps 6;

QY 226 SNGMW-----FSGRD-----GDLINASNAGTVIOADHN-----MDGASIVIOHTNGFVSSY 271
Db 83 AGGMMAHKHSQDDPAVPSGTEVLAAHGCTVVKAGNGAGDPAGNAIVIKHNGTYSQY 142
QY 272 IHIDAQVKTGDTVTRTGORIAASMKN--QPSGAAL-FEPRISHN-GVYVDPPLTVLK 322
Db 143 AHLGRIDVRIGQVAVATGCHIALSGNTGNSGPHLHFEIRTPSYGSDVDPAFLR 197

RESULT 4

US-10-156-761-14779
; Sequence 14779, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: HORIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 14779
; LENGTH: 397
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-14779

Query Match 7.6%; Score 127.5; DB 9; Length 397;
Best Local Similarity 26.5%; Pred. No. 0.0291;
Matches 79; Conservative 23; Mismatches 131; Indels 65; Gaps 14;

QY 43 SSGSHRTSGSGGLAIGS---QVITDSC-GVPRRYOVKQDPTVSKIA-----QRY 88
Db 141 SSGSHRTADRSYKRYADVAPQTPPOSQAGTAMETVVLHGDFLISGIADSRHVGWGEOLY 200
QY 89 GLMWRREIGHINNINSYTTTGTGMLTMSGDLKVRERSISSGVNTAHTPSPAVAVOSSRPP 148
Db 201 AANRGITGADPD-----ILPGORLIRA---KATRTIDASTST-HKSTSSSKSSTK 251
QY 149 VQOPAFVOKPTPPVVVVKKPTPPPVVQOPAPVAPVTEAPATSSGVMQFRYPVAGATN 208
Db 252 ASSDR-----TERATTSLSLV---APVNAAGTGYAHAGSS--WSKGYHTGVDF 295

[illegible]

```

RESULT 5
US-09-468-147-166
: Sequence 166, Application US/09468147A
: Publication No. US20030049601A1
: GENERAL INFORMATION:
: APPLICANT: Abbott Laboratories
: APPLICANT: Schlauder, George G.
: APPLICANT: Erker, James C.
: APPLICANT: Desai, Suresh M.
: APPLICANT: Dawson, George J.
: APPLICANT: Mushawar, I. K.
: TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTING
: TITLE OF INVENTION: HEPATITIS E VIRUS
: FILE REFERENCE: 6232.US..P1
: CURRENT APPLICATION NUMBER: US/09/468,147A
: CURRENT FILING DATE: 1999-12-21
: EARLIER APPLICATION NUMBER: US 09/173,141
: EARLIER FILING DATE: 1998-10-15
: EARLIER APPLICATION NUMBER: US 60/061,199
: EARLIER FILING DATE: 1997-10-15
: NUMBER OF SEQ. ID NOS: 258
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ. ID NO 166
: LENGTH: 1708
: TYPE: PRT
: ORGANISM: Hepatitis E Virus
: FEATURE:
: OTHER INFORMATION: Xaa = Unknown or Other at position 322
: OTHER INFORMATION: Xaa = Unknown or Other at position 331
: OTHER INFORMATION: Xaa = Unknown or Other at position 445
: OTHER INFORMATION: Xaa = Unknown or Other at position 448
: OTHER INFORMATION: Xaa = Unknown or Other at position 634
: OTHER INFORMATION: Xaa = Unknown or Other at position 646
: OTHER INFORMATION: Xaa = Unknown or Other at position 811
: OTHER INFORMATION: Xaa = Unknown or Other at position 1553
: OTHER INFORMATION: Xaa = Unknown or Other at position 1578
: OTHER INFORMATION: Xaa = Unknown or Other at position 1691
: US-09-468-147-166

```

	Query Match	7.6%	Score 127 ; DB 9 ; Length 1708 ;
	Best Local Similarity	22.7% ; Pred No. 0.19 ;	
	Matches	80 ; Conservative 37 ; Mismatches 120 ; Indels 116 ; Gaps 19	
Oy	27 TCIAGC---ASKPTYNSTGSGSHRTSGSGLAI-----GSQYITDSSGVENRYOKV 76		
Db	574 TTIVDGAILHANGPEEEVLISFDASQSMGASHSLTYELTPALQLVKISSNIGDTCATPP 633		
Oy	77 QGDPTS-----KIAORYGLN--WR-----EIGHI-----NN 100		
Db	634 XCGAPSAAPGEVXACSAALRYNRKTQRHSLTGMLMHEGLLGTFPPSPGHINESANP 693		
Oy	101 LNSSTTIYTGGWLLT--WGDLKVRRERSISSGVNTAHTPSFVAVQSSRPV-----149		
Db	694 FCGEGTLTRYMTWSTGFSFSDFPAPAAPASAABGLPFPT-----PVSVDIWYLPPPS 747		
Oy	150 -QQH-----PAVKCP---TPVVVVKKKPTPPPVQCAPVAPPYTEAFATSGSGWQC 199		
Db	748 EESHADASVSVPEPPAGLTSPIVLT--PPPPPVPVRKAATSPPPTRR-----LL 796		
Oy	200 FRYPVGATNPVVRRCGTATVAGSYTYSNGMWTSGRDDGLINASNACTYIQADHNMDGASI 259		
Db	797 YTPPGGA-----KYTAGSLKESDDW-----LVNASNPQ-----HRPGG- 831		

```

QY      260 VIQHTNGFVSSYIHIKDAQVKTGDIYRTGQRATSMKNQ-PGSAALFEERISRN 311
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      832 -LCH----AFYQRFPEAPVSTEFIMREGLAAYTLTPRIPIHAAVAPDYRVEQN 878

```

```

RESULT 6
US-09-738-626--6241
: Sequence 6241, Application US/09738626
: Publication No. US20020197605A1
GENERAL INFORMATION:
APPLICANT: NAKAGAWA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: OCHIAI, KEIKO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, MOKO
APPLICANT: SENOH, AKIHIRO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738, 626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 6241
LENGTH: 205
TYPE: FRT
: ORGANISM: Corynebacterium glutamicum
US-09-738-626--6241

```

	Query Match	7.5%	Score 125.5;	DB 9;	Length 205;
	Best Local Similarity	33.8%;	Pred. No. 0.018;		
	Matches	52;	Conservative	14;	Mismatches 73; Indels 15; Gaps 5;
QY	179 APVAPPTAEAFPATGSSGVNQFRYPVG-ATNPVRREFGTAVGSTVTS-----	226			
Db	41 ATMAPASAQIDYDALSSGVADTVEAAGVATTAAAPATAVARPANGFTTSGFGRWGFH	100			
QY	227 NGMVFSGDEDLINASNAGTVIQADHNMD-GASTIVIOHTNGFVSSYYHIKDAQYKTGDTV	285			
Db	101 NGDIANSIGTPITVAWNAVGTVISSEGPASGVQWMIRIHDDGSISITYGMETLYVSVEREV	160			
QY	286 RTGORIASMNKP-SGAALFEFRSRSGVY-VDP	317			
Db	161 AAGEIAGMGSGQFSSTGSHLHEITHPDGVVPVD	194			

```

RESULT 7
US-10-060-036-71
; Sequence 71. Application US/10060036
; Publication No. US20030073144A1
; GENERAL INFORMATION:
; APPLICANT: Benson, Darin R.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Persing, David H.
; APPLICANT: Hepler, William T.
; APPLICANT: Jiang, Yugu
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; OF INTRACRANIAL AND DIAGNOSIS OF PANCREATIC CANCER
; FILE REFERENCE: 210121.566
; CURRENT APPLICATION NUMBER: US/10/060.036
; CURRENT FILING DATE: 2002-01-30
; NUMBER OF SEQ ID NOS: 4560
; SOFTWARE: FastSeq for Windows Version 4.0

```

Wed Jul 9 10:00:59 2003

us-10-018-706-2.rapb

Page 4

SEQ ID NO 71
LENGTH: 1192
TYPE: PRT
ORGANISM: Homo sapiens
US-10-060-036-71

Query Match
Best Local Similarity 7.1%; Score 118; DB 9; Length 1192;
Matches 36; Conservative 8; Mismatches 33; Indels 14; Gaps 6;

QY 132 NTANTSPV---AVQSSRPVQOHPAVOKTPTPVVVK---KPTPTPVVQOPAPVAPV 185
DB 115 STVPASPLSAAVSPSKLPEDDEPPARPPPPASVSQAEPVWTPPA---PAPAAPPS 171
QY 186 T-EAPFATGSSGVWQ---FRYPVGTATNPVVR 212
DB 172 TPAAPKRRGSSGSVDETLFLALP-AASEPVIR 201

RESULT 8
US-09-789-386-2
Sequence 2, Application US/09789386
Patent No. US20020010324A1
GENERAL INFORMATION:
APPLICANT: MICHALOVICH, DAVID
APPLICANT: PRINJHA, RABINDER KUMAR
TITLE OF INVENTION: NOVEL COMPOUNDS
FILE REFERENCE: GP-30165-C1
CURRENT APPLICATION NUMBER: US/09/789,386
CURRENT FILING DATE: 2001-02-21
PRIOR FILING DATE: 1999-07-19
PRIOR APPLICATION NUMBER: U.K. 9916898.1
PRIOR FILING DATE: 1998-07-22
PRIOR APPLICATION NUMBER: U.K. 9816024.5
PRIOR FILING DATE: 1998-07-22
PRIOR APPLICATION NUMBER: US 09/359,208
NUMBER OF SEQ ID NOS: 6
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 2
LENGTH: 1192
TYPE: PRT
ORGANISM: HOMO SAPIENS
US-09-789-386-2

Query Match
Best Local Similarity 7.1%; Score 118; DB 10; Length 1192;
Matches 36; Conservative 8; Mismatches 33; Indels 14; Gaps 6;

QY 132 NTANTSPV---AVQSSRPVQOHPAVOKTPTPVVVK---KPTPTPVVQOPAPVAPV 185
DB 115 STVPASPLSAAVSPSKLPEDDEPPARPPPPASVSQAEPVWTPPA---PAPAAPPS 171
QY 186 T-EAPFATGSSGVWQ---FRYPVGTATNPVVR 212
DB 172 TPAAPKRRGSSGSVDETLFLALP-AASEPVIR 201

RESULT 9
US-09-758-140-6
Sequence 6, Application US/09758140
Patent No. US20020012965A1
GENERAL INFORMATION:
APPLICANT: Strittmatter, Stephen M.
TITLE OF INVENTION: No. US20020012965A1o Receptor-Mediated Blockade of Axonal Growth
FILE REFERENCE: 44574-5073-US
CURRENT APPLICATION NUMBER: US/09/758,140
CURRENT FILING DATE: 2001-01-12
PRIOR FILING DATE: US/09/758,140
PRIOR APPLICATION NUMBER: US 60/175,707
PRIOR FILING DATE: 2000-01-12
PRIOR APPLICATION NUMBER: US 60/207,366
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 60/236,378
PRIOR FILING DATE: 2000-09-29

NUMBER OF SEQ ID NOS: 20
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 6
LENGTH: 1192
TYPE: PRT
ORGANISM: Homo sapiens
US-09-758-140-6

Query Match
Best Local Similarity 7.1%; Score 118; DB 10; Length 1192;
Matches 36; Conservative 8; Mismatches 33; Indels 14; Gaps 6;

QY 132 NTANTSPV---AVQSSRPVQOHPAVOKTPTPVVVK---KPTPTPVVQOPAPVAPV 185
DB 115 STVPASPLSAAVSPSKLPEDDEPPARPPPPASVSQAEPVWTPPA---PAPAAPPS 171
QY 186 T-EAPFATGSSGVWQ---FRYPVGTATNPVVR 212
DB 172 TPAAPKRRGSSGSVDETLFLALP-AASEPVIR 201

RESULT 10
US-09-893-348-23
Sequence 23, Application US/09893348
Patent No. US20020072493A1
GENERAL INFORMATION:
APPLICANT: EISENBACH-SCHWARTZ, Michael
APPLICANT: COHEN, Itun R.
APPLICANT: BESERMAN, Pierre
APPLICANT: MOSONIGO, Alon
APPLICANT: MOLEME, Gila
TITLE OF INVENTION: ACTIVATED T-CELLS, NERVOUS SYSTEM-SPECIFIC ANTIGENS AND THEIR USES
FILE REFERENCE: EIS-SCHWARTZ-2A
CURRENT APPLICATION NUMBER: US/09/893,348
CURRENT FILING DATE: 2001-06-28
PRIOR FILING DATE: 1999-05-19
PRIOR APPLICATION NUMBER: US 09/314,161
PRIOR FILING DATE: 1999-05-19
PRIOR APPLICATION NUMBER: US 09/218,277
PRIOR FILING DATE: 1998-12-22
PRIOR APPLICATION NUMBER: PCT/US98/14715
PRIOR FILING DATE: 1998-07-21
PRIOR APPLICATION NUMBER: IL 124500
PRIOR FILING DATE: 1998-05-19
NUMBER OF SEQ ID NOS: 29
SOFTWARE: PatentIn version 3.1
SEQ ID NO 23
LENGTH: 1192
TYPE: PRT
ORGANISM: Homo sapiens
US-09-893-348-23

Query Match
Best Local Similarity 7.1%; Score 118; DB 10; Length 1192;
Matches 36; Conservative 8; Mismatches 33; Indels 14; Gaps 6;

QY 132 NTANTSPV---AVQSSRPVQOHPAVOKTPTPVVVK---KPTPTPVVQOPAPVAPV 185
DB 115 STVPASPLSAAVSPSKLPEDDEPPARPPPPASVSQAEPVWTPPA---PAPAAPPS 171
QY 186 T-EAPFATGSSGVWQ---FRYPVGTATNPVVR 212
DB 172 TPAAPKRRGSSGSVDETLFLALP-AASEPVIR 201

RESULT 11
US-09-972-599A-6
Sequence 6, Application US/09972599A
Patent No. US2002007295A1
GENERAL INFORMATION:
APPLICANT: STRITTMATTER, STEPHEN M.
TITLE OF INVENTION: NOGO-RECEPTOR-MEDIATED BLOCKADE OF AXONAL GROWTH
FILE REFERENCE: C077 CIP US
CURRENT APPLICATION NUMBER: US/09/972,599A

;; CURRENT FILING DATE: 2001-10-06
;; PRIOR APPLICATION NUMBER: PCT/US01/01041
;; PRIOR FILING DATE: 2001-01-12
;; PRIOR APPLICATION NUMBER: 09/758,140
;; PRIOR FILING DATE: 2001-01-12
;; PRIOR APPLICATION NUMBER: 60/236,378
;; PRIOR FILING DATE: 2000-09-29
;; PRIOR APPLICATION NUMBER: 60/207,366
;; PRIOR FILING DATE: 2000-05-26
;; PRIOR APPLICATION NUMBER: 60/175,707
;; PRIOR FILING DATE: 2000-01-12
;; NUMBER OF SEQ ID NOS: 57
;; SOFTWARE: PatentIn Ver. 2.1
;; SEQ ID NO 6
;; LENGTH: 1192
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-09-972-599A-6

Query Match 7.1%; Score 118; DB 10; Length 1192;
Best Local Similarity 39.6%; Pred. No. 0.62;
Matches 36; Conservative 8; Mismatches 33; Indels 14; Gaps 6;

QY 132 NTATPSPV--AVOSSRPVQOHAPAVQKTPPVVVK--KPTPTPVVQOAPVAPV 185
Db 115 STVPAPSPLSAAVSPSKLPEDDEPPARPPPPASVSPQAPVWTPPA--PAPAPPS 171
QY 186 T-EAPFATGSSGVQWQ--FRYPVATNPVVR 212
Db 172 TPAAPKRRGSSGVDETITFALP-AASEPVR 201

RESULT 12
US-09-738-626-4458
;; Sequence 4458, Application US/09738626
;; Publication No. US20020197605A1
;; GENERAL INFORMATION:
;; APPLICANT: NAKAGAWA, SATOSHI
;; APPLICANT: MIZOGUCHI, HIROSHI
;; APPLICANT: ANDO, SEIKO
;; APPLICANT: HAYASHI, MIKIRO
;; APPLICANT: OCHIAI, KEIKO
;; APPLICANT: YOKOI, HARUHIKO
;; APPLICANT: TATEISHI, NAOKO
;; APPLICANT: SENOH, AKIHIRO
;; APPLICANT: IKEDA, MASATO
;; APPLICANT: OZAKI, AKIO
;; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;; FILE REFERENCE: 249-125
;; CURRENT APPLICATION NUMBER: US/09/738, 626
;; CURRENT FILING DATE: 2000-12-18
;; PRIOR APPLICATION NUMBER: JP 99/377484
;; PRIOR FILING DATE: 1999-12-16
;; PRIOR APPLICATION NUMBER: JP 00/159162
;; PRIOR FILING DATE: 2000-04-07
;; PRIOR APPLICATION NUMBER: JP 00/280988
;; PRIOR FILING DATE: 2000-08-03
;; NUMBER OF SEQ ID NOS: 7059
;; SOFTWARE: PatentIn ver. 3.0
;; SEQ ID NO 4458
;; LENGTH: 237
;; TYPE: PRT
;; ORGANISM: Corynebacterium glutamicum
US-09-738-626-4458

Query Match 7.0%; Score 117; DB 9; Length 237;
Best Local Similarity 22.2%; Pred. No. 0.1;
Matches 70; Conservative 44; Mismatches 93; Indels 108; Gaps 15;
QY 10 QNCKPIKRLGIFVITTCILAGCASKPTYNSTGSGSHRSGGGLAIGQVITTDQGV 69
Db 13 QTSPTKRGVAFVAVAT-----GAVSTAGG-AVAAQ--ASNQPV 51

QY 70 PNRVQKQDPTVSKIAQRVG---LNRREIGHINNINSYTIYTGWLTLMSGDLKVRERS 126
Db 52 EVNFEILTANDTTLVAGSSAPQLSTIAEFKPVNL-----GDQYK--- 92
QY 127 ISSGVNTATPSPVAVQSSRPVQOHAPAVQKTPPVVVKKPTPTPVVQOAPVAPVT 186
Db 93 -----TTQYNADRIQADLDAGPS-----VVRPA----- 116
QY 187 EAPFATGSSGVQWQFRYPVATNPVVRFGATVAGSTVTSNGMFFSGRQDGLINASAGT 246
Db 117 EGSYTSG-----FGA-----RWGT-----NHNQVDIANATGPILANDGT 152
QY 247 VIQADHNM-DGASIVIOHTNGFPVSYIHKDAQVKTGTVRTQRIASMRNQ--PSGAL 303
Db 153 VIDAGPAGSGRMVRRIQHEGDTITVYGHMETVETVYQGVKAGERIAGMSRSGSTSHL 212
QY 304 -FEFRISRNGVYDP 317
Db 213 HFEVYPAGGGA-VDP 226

RESULT 13
US-10-155-400-1
;; Sequence 1, Application US/10155400
;; Publication No. US20030108988A1
;; GENERAL INFORMATION:
;; APPLICANT: DING, SHI-YOU
;; APPLICANT: ADNEY, WILLIAM S.
;; APPLICANT: VINZANT, TODD B.
;; APPLICANT: HIMMEL, MICHAEL E.
;; TITLE OF INVENTION: THERMAL TOLERANT AVICELASE FROM ACIDOTHERMUS
;; FILE REFERENCE: NREL 01-36A
;; CURRENT APPLICATION NUMBER: US/10/155,400
;; CURRENT FILING DATE: 2002-10-22
;; NUMBER OF SEQ ID NOS: 7
;; SOFTWARE: PatentIn Ver. 2.1
;; SEQ ID NO 1
;; LENGTH: 957
;; TYPE: PRT
;; ORGANISM: Acidothermus cellulolyticus
;; FEATURE:
;; NAME/KEY: MOD RES
;; LOCATION: (957)
;; OTHER INFORMATION: Any amino acid
US-10-155-400-1

Query Match 6.7%; Score 112; DB 9; Length 957;
Best Local Similarity 24.5%; Pred. No. 1.4;
Matches 67; Conservative 38; Mismatches 104; Indels 64; Gaps 15;

QY 34 ASKPTYNSTGSGSHR-----TSGSGGLAIGSQ-----YITDSQGVPNRYQVQGD 79
Db 677 ASGLYHSTWGGSSWALITGVSSAVNVGFGKSAAGSSYPVAFVGTIGTGYARRSDCG 736
QY 80 T---VSKIAQRGLMWRRE--IGHINNINSYTIYTGWLTLMSGDL----- 120
Db 737 TTVWLINDDOHQYG-NMGQAITGDHANLRVYICNIGRGIVY--GDIGCAPSSSPERSVS 793
QY 121 KVRERSISGVNTATPSPVAVQSSRPVQOHAPAVQKTPPVVVKKPTPTPVVQOAP 178
Db 794 PSASPISPPSPSSSPSPSPSSSPSSSPSSSPSSSPSSSPSSSPSSSPSSSPSSSP 850
QY 179 APVAPVYTE--APFATGSSGVQWQFRYPVATNP-----VVRFGTATVAGSTVTS 226
Db 851 SPSSSPSSSPSPSPSSSPSSSPSSSPSSSPSSSPSSSPSSSPSSSPSSSPSSSPSS 910
QY 227 NGMFFSGRQDGLINASAGTVIDADHNMDCASI 259
Db 911 R-YWFT-RDG-----GSSITLV--YNCDMAAI 932

RESULT 14

```
US-09-917-384-1
; Sequence 1, Application US/09917384
; Publication No. US20030096342A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: DECKER, STEPHEN R.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS
; FILE REFERENCE: 40170.6US01
; CURRENT APPLICATION NUMBER: US/09/917,384
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1228
; TYPE: PRT
; ORGANISM: Artificial Sequence
; OTHER INFORMATION: Description of Artificial Sequence: Segment of
US-09-917-384-1
Query Match
Best Local Similarity 24.3%; Score 112; DB 9; Length 1228;
Matches 65; Conservative 20; Mismatches 92; Indels 90; Gaps 12;
QY 51 SSGGLAIGSOVITDSOGVPRYOVKQGDVTSKIAQRYGLNWEIGH-----INNIN 102
DB 992 SRGVOPEGSQTATGVTAGHTMNWQGOQTS-----WKIISYVLTGATGISNLD 1042
QY 103 -----SSYTIYTGQWLTLMSGDLKVERSISSGVNTAHTPSPAVQSSR 146
DB 1043 LKAIFADAAARGSLNSTDLDVEAGFEIWMQGGGLGSNFSVSV-TSGTSSPTSPSPPT 1101
QY 147 PPVOQHAPVQKPTPPVVVKKPTPPVVOQAPAPVAPVTEAPATGSSGVMOFRYPVGA 206
DB 1102 P-----TPSP-----TPSPSPPTSPSPSPSP-----SSSGV-----ACRA 1133
QY 207 TNPVRRFGTATVAGSTVTSNG-----MFGSGRDDLINASNAGTVIOADHNMDGA 257
DB 1134 TYVNSDMGSGFTATVTNTGSRATNGWTAMSGNGQTNTNMTAL-----TOSGA 1187
QY 258 SI-----VIO-----HTNGFVSSY 271
DB 1188 SVTATNLSYNNVIOPGOSTTFGFGNGSY 1214
RESULT 15
US-09-917-383-1
; Sequence 1, Application US/09917383
; Publication No. US20030104522A1
; GENERAL INFORMATION:
; APPLICANT: DING, SHI-YOU
; APPLICANT: ADNEY, WILLIAM S.
; APPLICANT: VINZANT, TODD B.
; APPLICANT: DECKER, STEPHEN R.
; APPLICANT: HIMMEL, MICHAEL E.
; TITLE OF INVENTION: THERMAL TOLERANT CELLULOSE FROM ACIDOTHERMUS
; FILE REFERENCE: 40170.6US01
; CURRENT APPLICATION NUMBER: US/09/917,383
; CURRENT FILING DATE: 2001-07-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1228
; TYPE: PRT
; ORGANISM: Artificial Sequence
; OTHER INFORMATION: Description of Artificial Sequence: Segment of
```

```
OTHER INFORMATION: GUXA
US-09-917-383-1
```

```
Query Match
Best Local Similarity 24.3%; Score 112; DB 9; Length 1228;
Matches 65; Conservative 20; Mismatches 92; Indels 90; Gaps 12;
```

```
QY 51 SSGGLAIGSOVITDSOGVPRYOVKQGDVTSKIAQRYGLNWEIGH-----INNIN 102
DB 992 SRGVOPEGSQTATGVTAGHTMNWQGOQTS-----WKIISYVLTGATGISNLD 1042
QY 103 -----SSYTIYTGQWLTLMSGDLKVERSISSGVNTAHTPSPAVQSSR 146
DB 1043 LKAIFADAAARGSLNSTDLDVEAGFEIWMQGGGLGSNFSVSV-TSGTSSPTSPSPPT 1101
QY 147 PPVOQHAPVQKPTPPVVVKKPTPPVVOQAPAPVAPVTEAPATGSSGVMOFRYPVGA 206
DB 1102 P-----TPSP-----TPSPSPPTSPSPSPSP-----SSSGV-----ACRA 1133
QY 207 TNPVRRFGTATVAGSTVTSNG-----MFGSGRDDLINASNAGTVIOADHNMDGA 257
DB 1134 TYVNSDMGSGFTATVTNTGSRATNGWTAMSGNGQTNTNMTAL-----TOSGA 1187
QY 258 SI-----VIO-----HTNGFVSSY 271
DB 1188 SVTATNLSYNNVIOPGOSTTFGFGNGSY 1214
```

```
RESULT 16
US-10-101-464A-930
; Sequence 930, Application US/10101464A
; Publication No. US20030046728A1
; GENERAL INFORMATION:
; APPLICANT: Strabala, Timothy
; APPLICANT: Nieuwenhuizen, Nicolaas
; TITLE OF INVENTION: Compositions Isolated from Plant Cells
; TITLE OF INVENTION: and their Use in the Modification of Plant Cell Signaling
; FILE REFERENCE: 11000.1020C2
; CURRENT APPLICATION NUMBER: US/10/101,464A
; CURRENT FILING DATE: 2002-03-18
; PRIOR APPLICATION NUMBER: 09/704,302
; PRIOR FILING DATE: 2000-11-01
; PRIOR APPLICATION NUMBER: 09/228,986
; PRIOR FILING DATE: 1999-01-12
; PRIOR APPLICATION NUMBER: 60/162,866
; PRIOR FILING DATE: 1999-11-01
; PRIOR APPLICATION NUMBER: PCT/US00/00724
; PRIOR FILING DATE: 2000-01-11
; NUMBER OF SEQ ID NOS: 989
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 930
; LENGTH: 312
; TYPE: PRT
; ORGANISM: Pinus radiata
US-10-101-464A-930
```

```
Query Match
Best Local Similarity 21.5%; Score 110.5; DB 9; Length 312;
Matches 52; Conservative 20; Mismatches 51; Indels 119; Gaps 9;
```

```
QY 35 SKPTVNSTSGSGSHRISGSGLAIGSOVITDSQ-GVPRYOVKQ-----DT- 80
DB 139 ANPLNPLTD-----LQIGVGVIPRICKSCSQAOVXNGTMTLITYVQPVDT 186
QY 81 -VSKIAQRYGLNWEIGHINNINSSYTIYTGQWLTLMSGDLKVERSISSGVNTAHTPS 139
DB 187 NLSSITRKFGSDLONFKSLNGMNSTLTAYT-----T 217
QY 140 VAVQSSRPVVOQHAPVQKPTPPVVVKKPTPPVVOQAPAPVAPVTEAPATGSSGVMO 199
DB 218 VLVVPSORPVLSQVSSSPPP-----PPPPP----- 245
```

Qy	200	FRYPVAGINPVVRRFGTAIVAGISTVTSNMMPSGDRDILINASNAGTIVIQADHMDGASI	259
		: :	
Db	246	-----AATSGNNTTSGG-----GSKNGVYIGR--SLGGAVA	274
Qy	260	VI 261	
		::	
Db	275	LV 276	

```

RESULT 17
US-10-156-761-9957
: Sequence 9957, Application US/10156761
: Publication No. US20030119018A1
: GENERAL INFORMATION:
: APPLICANT: OMURA, SATOSHI
: APPLICANT: IKEDA, HARUO
: APPLICANT: ISHIKAWA, JUN
: APPLICANT: HORIKAWA, HIROSHI
: APPLICANT: SHIBA, TADAYOSHI
: APPLICANT: SAKAKI, YOSHIYUKI
: APPLICANT: HATTORI, MASAHIRA
: TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
: FILE REFERENCE: 249-262
: CURRENT FILING DATE: 2002-05-29
: PRIOR FILING DATE: 2002-05-29
: PRIOR APPLICATION NUMBER: JP 2001-204089
: PRIOR FILING DATE: 2001-05-30
: PRIOR APPLICATION NUMBER: JP 2001-272697
: PRIOR FILING DATE: 2001-08-02
: NUMBER OF SEQ ID NOS: 15109
: SEQ ID NO 9957
: LENGTH: 258
: TYPE: PRT
: ORGANISM: Streptomyces avermitilis
US-10-156-761-9957

```

Query Match	6.6%	Score 110;	DB 9;	Length 258;
Best Local Similarity	25.2%	Pred. No. 0.4;	Mismatches 108;	Indels 26;
Matches	55;	Conservative	29;	Gaps 8

QY	131	VNTAHTSPVAVQSSRRPVPQOHAPAVQKPTPPVVKKPTPTTPPVVQOPAPVAPVTEAF	190
Db	35	IGTLAAPALPAEPAPAEPEVEQTGLTQAVSICDSVAQYDACAQVQMAAEAAAKKAAEE	94
QY	191	ATGSSGVWQFPPVPGATNPVVR-----RFGT--AAVASTVTS-----NGMFFSG-----	233
Db	95	AARKKAELAEKEHAAKAPAAAEERKELNTFVADTASVSTGYKSGSLSSSGHTGV	154
QY	234	----RDGDLINASNAGTIVQAD--HNMDASITVQHTNGFVSSYIHKDAQVKTGDTVRG	288
Db	155	DFHAASGTSVHAVSGTVAEAGMGCAYGNNIYTKNMDGYTNGYGHLSISIGSVGQTVTPG	214
QY	289	QR--IASMKQPSGAAI--FEFRISKN--GVYDDELTVLK	322
Db	215	QOIGISGATGNTTGPHLFEARTTAAYESDDIDPVAYLR	252

RESULT 18
US-09-925-301-1116
Sequence 1116, Application US/09925301
Patent No. US20020052308A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OR INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PAL06
CURRENT APPLICATION NUMBER: US/09/925,301
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05862
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 1694
SOFTWARE: PatentIn Ver. 2.0

```

? SEQ ID NO 1116
? LENGTH: 360
? TYPE: PRT
? ORGANISM: Homo sapiens
? FEATURE:
? NAME/KEY: SITE
? LOCATION: (5)
? OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
? NAME/KEY: SITE
? LOCATION: (19)
? OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
? NAME/KEY: SITE
? LOCATION: (29)
? OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
? NAME/KEY: SITE
? LOCATION: (38)
? OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
? JS-09-925-301-1116

```

	Query Match	6.6%;	Score 110;	DB 10;	Length 360;	
	Best Local Similarity	22.0%;	Pred. No.61;			
	Matches	51;	Conservative	21;	Mismatches	80; Indels 80; Gaps 7
Qy	9	SONOKPIKRLGLLIGVITTCILACGASKPYNSTSGSGSHRTSGGGIALISQVITDSQG	68			
		:::::		:::::		
Db	110	SINORPETHSG-----SSGGSSRENGSSSIGIPIAVPT----	144			
Qy	69	VPNKYVKQGDGTSCIAQRVGLNMRBEIGHINNLSYTYTGQWLTLMSGDLKXRERSIS	128			
		:::::		:::::		
Db	145	-PSPPITGGPAAPGAPGSQYGTMTROISRHNSTTS-----STS	181			
Qy	129	SGVTAAHTPSPVAOVASRPVVOQH-----AVQKPTRPVV-----	163			
		:::::		:::::		
Db	182	SG-GYRRTSPSVTAQFSAQPHVNGGPLXSQNSISIAPEPPMPQLTPQPLTGFVARVQEN	240			
Qy	164	VKKETPTTP-----VVQCAPAAPPTTEAPFRATGSGGNQWQFRIPVATNP	209			
		:::::		:::::		
Db	241	IADSPTRPPPPPDIMFDSDSPPPPPPVVDYEDEEAVALYQNDPVADGD	292			

```

RESULT 19
US-09-738-626-3548
: Sequence 3548, Application US/09738626
: Publication NO. US20020197605A1
: GENERAL INFORMATION:
: APPLICANT: NAKAGAWA, SATOSHI
: APPLICANT: MIZOGUCHI, HIROSHI
: APPLICANT: ANDO, SEIKO
: APPLICANT: HAYASHI, MIKIRO
: APPLICANT: OCHIAI, KEIKO
: APPLICANT: YOKOI, HARUHIKO
: APPLICANT: TATEISHI, NAOKO
: APPLICANT: SENOH, AKIHIRO
: APPLICANT: IKEDA, MASATO
: APPLICANT: OZAKI, AKIO
: TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
: FILE REFERENCE: 249-125
: CURRENT APPLICATION NUMBER: US/09/738, 626
: CURRENT FILING DATE: 2000-12-18
: PRIOR APPLICATION NUMBER: JP 99/377484
: PRIOR FILING DATE: 1999-12-16
: PRIOR APPLICATION NUMBER: JP 00/159162
: PRIOR FILING DATE: 2000-04-07
: PRIOR APPLICATION NUMBER: JP 00/280988
: PRIOR FILING DATE: 2000-08-03
: NUMBER OF SEQ ID NOS: 7059
: SOFTWARE: PatentIn ver. 3.0
: SEQ ID NO 3548
: LENGTH: 474
: TYPE: PRT
: ORGANISM: Corynebacterium glutamicum
US-09-738-626-3548

```

Wed Jul 9 10:00:59 2003

us-10-018-706-2.rapb

Page 8

Query Match 6.5%; Score 108.5; DB 9; Length 474;
Best Local Similarity 20.3%; Pred. No. 1.1;
Matches 70; Conservative 49; Mismatches 123; Indels 103; Gaps 14;
49 RTSGSGGLAIGSQVIT-----DSQGVNRYQVKO-----GPTVSKIAQRYGLNREIGHNN 100
48 KSTPRQDQISTGCGVLAESSQDDQGYORSYITNPTAVPVYGLSDVYGAAGLEIGYNI 107
101 LN-SSTIYTGQWLTLMGDLKXREISISGVTATPSPVAVQSSRPVQCH-----P 153
108 LNSGDSLSFTSCQLDVISG-----SPTGANIELTDPNMQCTIYEQLSOSGTYGAVY 160
154 AVCKPPTPVV-VYKRPPTPVVQOPAPAPVPTETPATGSGVQWQFRIYVQATNPVVR 212
161 ALRPSGCVLMASSSYDNDQIVDPATYADAMAE---YISTEGAPLNMHATOESLPPQS 217
213 RFGTAVAGSTVTSNGMWSGRDGLINASNACTVIQADHNDGASIVIQHTNGFVSSY- 271
218 IFKITTTAA-----LENGYSADSTVTA-----AAVTLPGTNTTLTNYG 257
272 -----HIKDAQVKTG-----DTVR-----TGORIA-SMKQOPSG- 300
258 GGTGAGGTTTLTLTAFOQLSCNTAFVETGIDVGADALRASAEDPGVQVTSLGIDNVGGL 317
301 -----AALFEFRISRNGVYVDPVTV 320
318 GEIPDDALGQSSIGORDVQMVNLQAAMVAGTVSNGVMEPYLV 362

RESULT 20
US-09-932-183A-2
Sequence 2, Application US/09932183A
Patent No. US20020127641A1
GENERAL INFORMATION:
APPLICANT: Estell, David A.
TITLE OF INVENTION: Proteases From Gram-Positive Organisms
FILE REFERENCE: GC394C1-US
CURRENT APPLICATION NUMBER: US/09/932,183A
CURRENT FILING DATE: 2001-08-17
PRIOR APPLICATION NUMBER: US 09/308,375
PRIOR FILING DATE: 1998-05-14
PRIOR APPLICATION NUMBER: PCT/US98/18828
PRIOR FILING DATE: 1998-09-08
PRIOR APPLICATION NUMBER: EP9719636.4
PRIOR FILING DATE: 1997-09-15
NUMBER OF SEQ ID NOS: 3
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 2
LENGTH: 2285
TYPE: PRT
ORGANISM: Bacillus subtilis
US-09-932-183A-2

Query Match 6.5%; Score 108.5; DB 10; Length 2285;
Best Local Similarity 20.3%; Pred. No. 7.6;
Matches 63; Conservative 51; Mismatches 149; Indels 47; Gaps 11;

22 FEVITTCILAGCASKPTVNSTSGSGSHRTSGSG-----ATGSOVITDSQGVNRYQVK 76
1398 YGIVISTTSSGTPSTGSGSYGKYSYNSAASKYNDVPALIAVIOQESGF--NAKAR 1455
77 QODTVSKIAQRYGLNREIGHNNLSY-----TIYTGQWLTLMGDLKXREISISG 130
1456 SGVAGMGLMQLMPATKSG-VNNNAVDPYQNMGGTKYLAQLEFGANVEALAAVYNG 1514
131 VNTATPSPVAVQSSRPVQCHPAPVQKPPVPPVAVVAKPPVPPVQOPAPVAPVTEAPF 190
1515 -----FQNVITRYGGLPPEKQNTYKIMANYSSLSATIST-----ASY 1555
191 ATGSSGVQWQFRIYVQATNPVVRFGTAVAGSTVTSNGMWSGRDGLINASNACTVIOA 250
1556 YNNNSA--FR-----VSSKYQGOESGLRSSPHKGTDFAAKAGTAKISLQSKVOIA 1604

251 DHNMD-GASIVTQHTNGFVSSYIH-IKDAQVKTGDTVTRTGORTASMKNOPSGAA-LFEFR 307
1605 GYSTAGKWWVVIKQDDGVAKYVMMLMTPSVKAGQSVKAGQTLGKVGSTGNSGNHLHQ 1664
308 ISRNGVYVDP 317
1665 IEQNGKITDP 1674

RESULT 21
US-10-245-103-90
Sequence 90, Application US/10245103
Publication No. US20030068778A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin
APPLICANT: Eaton, Dan
APPLICANT: Filvaroff, Ellen
APPLICANT: Goddard, Audrey
APPLICANT: Grimaldi, J. Christopher
APPLICANT: Gurney, Austin
APPLICANT: Smith, Victoria
APPLICANT: Stephan, Jean-Phillippe
APPLICANT: Wacande, Colin
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin
APPLICANT: Fong, Sherman
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3630R1C112
CURRENT APPLICATION NUMBER: US/10/245,103
CURRENT FILING DATE: 2002-09-17
PRIOR APPLICATION NUMBER: 10/197942
PRIOR FILING DATE: 2002-07-18
PRIOR APPLICATION NUMBER: 60/059114
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/063046
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/065027
PRIOR FILING DATE: 1997-11-10
PRIOR APPLICATION NUMBER: 60/079689
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/086478
PRIOR FILING DATE: 1998-05-22
PRIOR APPLICATION NUMBER: 60/087607
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/089801
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/090557
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090689
PRIOR FILING DATE: 1998-06-25
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 116
SEQ ID NO 90
LENGTH: 877
TYPE: PRT
ORGANISM: Homo Sapien
US-10-245-103-90

Query Match 6.5%; Score 108; DB 9; Length 877;
Best Local Similarity 24.4%; Pred. No. 2.6;
Matches 64; Conservative 28; Mismatches 94; Indels 76; Gaps 16;

23 GVITTCILAGCASKPTVNSTSGSGSHRTSGSG-----GLAISQVITDSQGVNRYQVK 73
446 GADPICVAGRCLS-PGDDGLISG-RRPDGCGGDDSTCRVLSGN--LTD-RGSPLEG 500
74 Q-----VKQDPTVSKIAQRYGLNREIGHNNLSY-----TIYTGQWLT----- 113
501 QKILMTIPAGALRLQIQTL-----RPSNVTALRGPGRSIINGNAVAVDDPPGXY 548
114 -----TLMGDLKXREISISGVT--ATPSPVAV-----QSSRPVQCHPAPVQKPPVPPV 164

Db 549 RAGTFRVRRNRPREEKGESLSAEGPTTQVDVYMIFOENPGVYQYVYISSPP---I 605
Qy 165 VKKPTPPVQV-QP--APVAPVTBAPATGSSGVMQF-----RYPGATN 208
Db 606 LENTPEPPVQLOPELIRPEPLAPRPARPTGTLQROVRIPQMAPPHPTPLGSPA 665
Qy 209 PVVRRFGTATVAGSTVTSNGMW 230
Db 666 AYWKRVGHSACSAS--CGKGVM 685

RESULT 22
US-10-245-107-90

/ Sequence 90, Application US/10245107
/ Publication No. US20030068779A1
/ GENERAL INFORMATION:
/ APPLICANT: Baker, Kevin
/ APPLICANT: Baton, Dan
/ APPLICANT: Filvaroff, Ellen
/ APPLICANT: Goddard, Audrey
/ APPLICANT: Grimaldi, J. Christopher
/ APPLICANT: Gurney, Austin
/ APPLICANT: Smith, Victoria
/ APPLICANT: Stephan, Jean-Philippe
/ APPLICANT: Watanabe, Colin
/ APPLICANT: Wood, William
/ APPLICANT: Zhang, Zemin
/ APPLICANT: Fong, Sherman
/ TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
/ FILE REFERENCE: P3630R1C71
/ CURRENT APPLICATION NUMBER: US/10/245,107
/ PRIOR FILING DATE: 2002-09-16
/ PRIOR APPLICATION NUMBER: 10/197942
/ PRIOR FILING DATE: 2002-07-18
/ PRIOR APPLICATION NUMBER: 60/059114
/ PRIOR FILING DATE: 1997-09-17
/ PRIOR APPLICATION NUMBER: 60/063046
/ PRIOR FILING DATE: 1997-10-24
/ PRIOR APPLICATION NUMBER: 60/065027
/ PRIOR FILING DATE: 1997-11-10
/ PRIOR APPLICATION NUMBER: 60/079689
/ PRIOR FILING DATE: 1998-03-27
/ PRIOR APPLICATION NUMBER: 60/086478
/ PRIOR FILING DATE: 1998-05-22
/ PRIOR APPLICATION NUMBER: 60/087607
/ PRIOR FILING DATE: 1998-06-02
/ PRIOR APPLICATION NUMBER: 60/089801
/ PRIOR FILING DATE: 1998-06-18
/ PRIOR APPLICATION NUMBER: 60/090557
/ PRIOR FILING DATE: 1998-06-24
/ PRIOR APPLICATION NUMBER: 60/090689
/ PRIOR FILING DATE: 1998-06-25
/ Remaining Prior Application data removed - See File Wrapper or PALM.
/ NUMBER OF SEQ ID NOS: 116
/ SEQ ID NO 90
/ LENGTH: 877
/ TYPE: PRT
/ ORGANISM: Homo Sapien
/ US-10-245-107-90

Query Match 6.5%; Score 108; DB 9; Length 877;

Best Local Similarity 24.4%; Pred. No. 2.6; Mismatches 94; Indels 76; Gaps 16;

Matches 64; Conservative 28; Mismatches 94; Indels 76; Gaps 16;
Qy 23 GVTTCTILACASKPTVNSTSGSHRTSGS-----GLAISQVITDSQGVPNRY 73
Db 446 GADPITVAGRCIS-PCDDGILISG-RRPDGCGVCGDDSTCRIVSGN--LTD-RGSPLAGY 500
Qy 74 Q-----VKQDPTVSKIAQRYGLNWRREIGHINNINSY-----TITYQWL----- 113
Db 501 QKILWIPAGALRIQIADL-----RPSNLYALRGPGGRSLIINGNMAVDPGGSY 548

Qy 114 ----TMSGDLKYREKSISSGVNT-AHTSPVAV-----OSSRPVQOHPAVOKPTPPVVV 164
Db 549 RAGTFRVRRNRPREEKGESLSAEGPTTQVDVYMIFOENPGVYQYVYISSPP---I 605
Qy 165 VKKPTPPVQV-QP--APVAPVTBAPATGSSGVMQF-----RYPGATN 208
Db 606 LENTPEPPVQLOPELIRPEPLAPRPARPTGTLQROVRIPQMAPPHPTPLGSPA 665
Qy 209 PVVRRFGTATVAGSTVTSNGMW 230
Db 666 AYWKRVGHSACSAS--CGKGVM 685

RESULT 23
US-10-245-143-90

/ Sequence 90, Application US/10245143
/ Publication No. US20030068780A1
/ GENERAL INFORMATION:
/ APPLICANT: Baker, Kevin
/ APPLICANT: Baton, Dan
/ APPLICANT: Filvaroff, Ellen
/ APPLICANT: Goddard, Audrey
/ APPLICANT: Grimaldi, J. Christopher
/ APPLICANT: Gurney, Austin
/ APPLICANT: Smith, Victoria
/ APPLICANT: Stephan, Jean-Philippe
/ APPLICANT: Watanabe, Colin
/ APPLICANT: Wood, William
/ APPLICANT: Zhang, Zemin
/ APPLICANT: Fong, Sherman
/ TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
/ FILE REFERENCE: P3630R1C90
/ CURRENT APPLICATION NUMBER: US/10/245,143
/ PRIOR FILING DATE: 2002-09-16
/ PRIOR APPLICATION NUMBER: 10/197942
/ PRIOR FILING DATE: 2002-07-18
/ PRIOR APPLICATION NUMBER: 60/059114
/ PRIOR FILING DATE: 1997-09-17
/ PRIOR APPLICATION NUMBER: 60/063046
/ PRIOR FILING DATE: 1997-10-24
/ PRIOR APPLICATION NUMBER: 60/065027
/ PRIOR FILING DATE: 1997-11-10
/ PRIOR APPLICATION NUMBER: 60/079689
/ PRIOR FILING DATE: 1998-03-27
/ PRIOR APPLICATION NUMBER: 60/086478
/ PRIOR FILING DATE: 1998-05-22
/ PRIOR APPLICATION NUMBER: 60/087607
/ PRIOR FILING DATE: 1998-06-02
/ PRIOR APPLICATION NUMBER: 60/089801
/ PRIOR FILING DATE: 1998-06-18
/ PRIOR APPLICATION NUMBER: 60/090557
/ PRIOR FILING DATE: 1998-06-24
/ PRIOR APPLICATION NUMBER: 60/090689
/ PRIOR FILING DATE: 1998-06-25
/ Remaining Prior Application data removed - See File Wrapper or PALM.
/ NUMBER OF SEQ ID NOS: 116
/ SEQ ID NO 90
/ LENGTH: 877
/ TYPE: PRT
/ ORGANISM: Homo Sapien
/ US-10-245-143-90

Query Match 6.5%; Score 108; DB 9; Length 877;

Best Local Similarity 24.4%; Pred. No. 2.6; Mismatches 94; Indels 76; Gaps 16;

Matches 64; Conservative 28; Mismatches 94; Indels 76; Gaps 16;
Qy 23 GVTTCTILACASKPTVNSTSGSHRTSGS-----GLAISQVITDSQGVPNRY 73
Db 446 GADPITVAGRCIS-PCDDGILISG-RRPDGCGVCGDDSTCRIVSGN--LTD-RGSPLAGY 500
Qy 74 Q-----VKQDPTVSKIAQRYGLNWRREIGHINNINSY-----TITYQWL----- 113

Wed Jul 9 10:00:59 2003

us-10-018-706-2.rapb

Page 10

Db 501 OKILMIPAGALRIQIAOL-----RPSNLYALRGPGRSIINGNMAVDPGSGY 548
QY 114 ----TWSGDLKVERSSISGVNT-AHTSPVAV-----OSSRPVQCHPAVOKPTPPVV 164
Db 549 RAGGTVEFRVNRPRREGKESLSAEGPTTQPDVYVMTFOENPGVFYQVYISSPP---I 605
QY 165 VKKPTPTPPVVO-QP--APVAPVTEAPFATSSGVWOF-----RYPVQATN 208
Db 606 LENTPEEPVPOLOPEILINVEPLAPAPRPARPTGLQROVRIPOMPAPPHRTPLGSPA 665
QY 209 PVVRRFGTATVAGSTVTSNGMW 230
Db 666 AYMKRVGHSACSAS--CGKGVW 685

RESULT 24
US-10-245-771-90

/ Sequence 90, Application US/10245771
/ Publication No. US20030068781A1
/ GENERAL INFORMATION:
/ APPLICANT: Baker, Kevin
/ APPLICANT: Baton, Dan
/ APPLICANT: Filvaroff, Ellen
/ APPLICANT: Goddard, Audrey
/ APPLICANT: Grimaldi, J. Christopher
/ APPLICANT: Guiney, Austin
/ APPLICANT: Smith, Victoria
/ APPLICANT: Stephan, Jean-Philippe
/ APPLICANT: Watanabe, Colin
/ APPLICANT: Wood, William
/ APPLICANT: Zhang, Zemin
/ APPLICANT: Fong, Sherman
/ TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
/ FILE REFERENCE: P3630R1C98
/ CURRENT APPLICATION NUMBER: US/10/245,771
/ PRIOR FILING DATE: 2002-09-16
/ PRIOR APPLICATION NUMBER: 10/197942
/ PRIOR FILING DATE: 2002-07-18
/ PRIOR APPLICATION NUMBER: 60/059114
/ PRIOR FILING DATE: 1997-09-17
/ PRIOR APPLICATION NUMBER: 60/063046
/ PRIOR FILING DATE: 1997-10-24
/ PRIOR APPLICATION NUMBER: 60/065027
/ PRIOR FILING DATE: 1997-11-10
/ PRIOR APPLICATION NUMBER: 60/079689
/ PRIOR FILING DATE: 1998-03-27
/ PRIOR APPLICATION NUMBER: 60/086478
/ PRIOR FILING DATE: 1998-05-22
/ PRIOR APPLICATION NUMBER: 60/087607
/ PRIOR FILING DATE: 1998-06-02
/ PRIOR APPLICATION NUMBER: 60/089801
/ PRIOR FILING DATE: 1998-06-18
/ PRIOR APPLICATION NUMBER: 60/090557
/ PRIOR FILING DATE: 1998-06-24
/ PRIOR APPLICATION NUMBER: 60/090689
/ PRIOR FILING DATE: 1998-06-25
/ Remaining Prior Application data removed - See file Wrapper or PALM.
/ NUMBER OF SEQ ID NOS: 116
/ SEQ ID NO 90
/ LENGTH: 877
/ TYPE: PRT
/ ORGANISM: Homo Sapien
US-10-245-771-90

Query Match 6.5%; Score 108; DB 9; Length 877;
Best Local Similarity 24.4%; Pred. No. 2.6;
Matches 64; Conservative 28; Mismatches 94; Indels 76; Gaps 16;
QY 23 GVTTTCIIAGCASKPTYNSTSGSGSHRTSGSG-----GIAISGVITDSQGVPNRY 73
Db 446 GAPDVCVGRCLIS-PCGDBILISG--RRPDGCGVCGDDSTCRVSGN--LTD-RGGPIGY 500

QY 74 Q-----VKQGDIVSKIAORYGLNWRREIGHINNUNSSY-----TIYTGOWL----- 113
Db 501 OKILMIPAGALRIQIAOL-----RPSNLYALRGPGRSIINGNMAVDPGSGY 548
QY 114 ----TWSGDLKVERSSISGVNT-AHTSPVAV-----OSSRPVQCHPAVOKPTPPVV 164
Db 549 RAGGTVEFRVNRPRREGKESLSAEGPTTQPDVYVMTFOENPGVFYQVYISSPP---I 605
QY 165 VKKPTPTPPVVO-QP--APVAPVTEAPFATSSGVWOF-----RYPVQATN 208
Db 606 LENTPEEPVPOLOPEILINVEPLAPAPRPARPTGLQROVRIPOMPAPPHRTPLGSPA 665
QY 209 PVVRRFGTATVAGSTVTSNGMW 230
Db 666 AYMKRVGHSACSAS--CGKGVW 685

RESULT 25
US-10-245-851-90

/ Sequence 90, Application US/10245851
/ Publication No. US20030068782A1
/ GENERAL INFORMATION:
/ APPLICANT: Baker, Kevin
/ APPLICANT: Baton, Dan
/ APPLICANT: Filvaroff, Ellen
/ APPLICANT: Goddard, Audrey
/ APPLICANT: Grimaldi, J. Christopher
/ APPLICANT: Guiney, Austin
/ APPLICANT: Smith, Victoria
/ APPLICANT: Stephan, Jean-Philippe
/ APPLICANT: Watanabe, Colin
/ APPLICANT: Wood, William
/ APPLICANT: Zhang, Zemin
/ APPLICANT: Fong, Sherman
/ TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
/ FILE REFERENCE: P3630R1C93
/ CURRENT APPLICATION NUMBER: US/10/245,851
/ PRIOR FILING DATE: 2002-09-16
/ PRIOR APPLICATION NUMBER: 10/197942
/ PRIOR FILING DATE: 2002-07-18
/ PRIOR APPLICATION NUMBER: 60/059114
/ PRIOR FILING DATE: 1997-09-17
/ PRIOR APPLICATION NUMBER: 60/063046
/ PRIOR FILING DATE: 1997-10-24
/ PRIOR APPLICATION NUMBER: 60/065027
/ PRIOR FILING DATE: 1997-11-10
/ PRIOR APPLICATION NUMBER: 60/079689
/ PRIOR FILING DATE: 1998-03-27
/ PRIOR APPLICATION NUMBER: 60/086478
/ PRIOR FILING DATE: 1998-05-22
/ PRIOR APPLICATION NUMBER: 60/087607
/ PRIOR FILING DATE: 1998-06-02
/ PRIOR APPLICATION NUMBER: 60/089801
/ PRIOR FILING DATE: 1998-06-18
/ PRIOR APPLICATION NUMBER: 60/090557
/ PRIOR FILING DATE: 1998-06-24
/ PRIOR APPLICATION NUMBER: 60/090689
/ PRIOR FILING DATE: 1998-06-25
/ Remaining Prior Application data removed - See file Wrapper or PALM.
/ NUMBER OF SEQ ID NOS: 116
/ SEQ ID NO 90
/ LENGTH: 877
/ TYPE: PRT
/ ORGANISM: Homo Sapien
US-10-245-851-90

Query Match 6.5%; Score 108; DB 9; Length 877;
Best Local Similarity 24.4%; Pred. No. 2.6;
Matches 64; Conservative 28; Mismatches 94; Indels 76; Gaps 16;
QY 23 GVTTTCIIAGCASKPTYNSTSGSGSHRTSGSG-----GIAISGVITDSQGVPNRY 73

Db 446 GARDICVAGRCLs-PGDDGILSG-RRPDGCGVGGDDSTCRVSGN--LTD-RGGPLGY 500
Qy 74 Q-----VKQGPVTSKIAORYGLNMRREIGHINNLSY-----TITYQWL----- 113
Db 501 OKILMIPAGLRQIOL-----RPSNYLALRPGGRSIIINGMVAVDPPGSY 548
Qy 114 -----TLMSGDLKVERSISSGVNT-AHTSPVAV-----QSSRPVQOHPAVQKPTPPVV 164
Db 549 RAGTIVFRNRPREBKGESLSAEGPTTPVDVYMIQENPGVYQYVSSDP--I 605
Qy 165 VKKPTPPVQV-QP--APVAPVTEAPFATGSSGVNQF-----RYPVQATN 208
Db 606 LENTPEPPVPQIQPILRVEPPLAPAPRAPRTPTGLQVRIPQMPAPPHRPTPLGSPA 665
Qy 209 PVVRFGTATVAGSTVTSNGMW 230
Db 666 AYWKRVGHSACSAS--CGKGW 685

RESULT 26

US-10-245-883-90
; Sequence 90, Application US/10245883
; Publication No. US20030068783A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin
; APPLICANT: Baton, Dan
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Goddard, Audrey
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin
; APPLICANT: Smith, Victoria
; APPLICANT: Stephan, Jean-Phillippe
; APPLICANT: Watande, Colin
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; APPLICANT: Fong, Sherman
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3630R1C70
; CURRENT APPLICATION NUMBER: US/10/245, 883
; CURRENT FILING DATE: 2002-09-16
; PRIOR APPLICATION NUMBER: 10/197942
; PRIOR FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/059114
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/063046
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/065027
; PRIOR FILING DATE: 1997-11-10
; PRIOR APPLICATION NUMBER: 60/079689
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/086478
; PRIOR FILING DATE: 1998-05-22
; PRIOR APPLICATION NUMBER: 60/087607
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/089801
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090689
; PRIOR FILING DATE: 1998-06-25
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 116
; SEQ ID NO 90
; LENGTH: 877
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-245-883-90

Query Match 6.5%; Score 108; DB 9; Length 877;
Best Local Similarity 24.4%; Pred. No. 2.6;
Matches 64; Conservative 28; Mismatches 94; Indels 76; Gaps 16;

Qy 23 GVITTCILACAKRPTYNSTSGSHRTSGG-----GLAIGQVITDSQGVNRY 73
Db 446 GARDICVAGRCLs-PGDDGILSG-RRPDGCGVGGDDSTCRVSGN--LTD-RGGPLGY 500
Qy 74 Q-----VKQGPVTSKIAORYGLNMRREIGHINNLSY-----TITYQWL----- 113
Db 501 OKILMIPAGLRQIOL-----RPSNYLALRPGGRSIIINGMVAVDPPGSY 548
Qy 114 -----TLMSGDLKVERSISSGVNT-AHTSPVAV-----QSSRPVQOHPAVQKPTPPVV 164
Db 549 RAGTIVFRNRPREBKGESLSAEGPTTPVDVYMIQENPGVYQYVSSDP--I 605
Qy 165 VKKPTPPVQV-QP--APVAPVTEAPFATGSSGVNQF-----RYPVQATN 208
Db 606 LENTPEPPVPQIQPILRVEPPLAPAPRAPRTPTGLQVRIPQMPAPPHRPTPLGSPA 665
Qy 209 PVVRFGTATVAGSTVTSNGMW 230
Db 666 AYWKRVGHSACSAS--CGKGW 685

RESULT 27

US-10-237-535-90
; Sequence 90, Application US/10237535
; Publication No. US20030073188A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin
; APPLICANT: Baton, Dan
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Goddard, Audrey
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin
; APPLICANT: Smith, Victoria
; APPLICANT: Stephan, Jean-Phillippe
; APPLICANT: Watande, Colin
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; APPLICANT: Fong, Sherman
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3630R1C3
; CURRENT APPLICATION NUMBER: US/10/237, 535
; CURRENT FILING DATE: 2002-09-06
; PRIOR APPLICATION NUMBER: 10/197942
; PRIOR FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/059114
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/063046
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/065027
; PRIOR FILING DATE: 1997-11-10
; PRIOR APPLICATION NUMBER: 60/079689
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/086478
; PRIOR FILING DATE: 1998-05-22
; PRIOR APPLICATION NUMBER: 60/087607
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/089801
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090689
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/091358
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091978
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/099803
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/106932
; PRIOR FILING DATE: 1998-11-03
; PRIOR APPLICATION NUMBER: 60/115554
; PRIOR FILING DATE: 1999-01-12

Wed Jul 9 10:00:59 2003

us-10-018-706-2.rapb

Page 12

PRIOR APPLICATION NUMBER: 60/119342
PRIOR FILING DATE: 1999-02-09
PRIOR APPLICATION NUMBER: 60/123957
PRIOR FILING DATE: 1999-03-12
PRIOR APPLICATION NUMBER: 60/123972
PRIOR FILING DATE: 1999-03-11
PRIOR APPLICATION NUMBER: 60/127372
PRIOR FILING DATE: 1999-04-01
PRIOR APPLICATION NUMBER: 60/131271
PRIOR FILING DATE: 1999-04-27
PRIOR APPLICATION NUMBER: 60/133459
PRIOR FILING DATE: 1999-05-11
PRIOR APPLICATION NUMBER: 60/135725
PRIOR FILING DATE: 1999-05-25
PRIOR APPLICATION NUMBER: 60/135729
PRIOR FILING DATE: 1999-05-25
PRIOR APPLICATION NUMBER: 60/135750
PRIOR FILING DATE: 1999-05-25
PRIOR APPLICATION NUMBER: 60/138385
PRIOR FILING DATE: 1999-06-09
PRIOR APPLICATION NUMBER: 60/140653
PRIOR FILING DATE: 1999-06-22
PRIOR APPLICATION NUMBER: 60/141037
PRIOR FILING DATE: 1999-06-23
PRIOR APPLICATION NUMBER: 60/144732
PRIOR FILING DATE: 1999-07-20
PRIOR APPLICATION NUMBER: 60/144758
PRIOR FILING DATE: 1999-07-20
PRIOR APPLICATION NUMBER: 60/144790
PRIOR FILING DATE: 1999-07-20
PRIOR APPLICATION NUMBER: 60/145228
PRIOR FILING DATE: 1999-07-20
PRIOR APPLICATION NUMBER: 60/145698
PRIOR FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: 60/146222
PRIOR FILING DATE: 1999-07-28
PRIOR APPLICATION NUMBER: 60/146843
PRIOR FILING DATE: 1999-08-03
PRIOR APPLICATION NUMBER: 60/148188
PRIOR FILING DATE: 1999-08-10
PRIOR APPLICATION NUMBER: 60/148513
PRIOR FILING DATE: 1999-08-12
PRIOR APPLICATION NUMBER: 60/149327
PRIOR FILING DATE: 1999-08-17
PRIOR APPLICATION NUMBER: 60/149395
PRIOR FILING DATE: 1999-08-17
PRIOR APPLICATION NUMBER: 60/150114
PRIOR FILING DATE: 1999-08-20
PRIOR APPLICATION NUMBER: 60/151700
PRIOR FILING DATE: 1999-08-31
PRIOR APPLICATION NUMBER: 60/151734
PRIOR FILING DATE: 1999-08-31
PRIOR APPLICATION NUMBER: 60/162506
PRIOR FILING DATE: 1999-10-29
PRIOR APPLICATION NUMBER: 60/170262
PRIOR FILING DATE: 1999-12-09
PRIOR APPLICATION NUMBER: 60/177118
PRIOR FILING DATE: 2000-01-20
PRIOR APPLICATION NUMBER: 60/179851
PRIOR FILING DATE: 2000-02-02
PRIOR APPLICATION NUMBER: 60/180921
PRIOR FILING DATE: 2000-02-08
PRIOR APPLICATION NUMBER: 60/187202
PRIOR FILING DATE: 2000-03-03
PRIOR APPLICATION NUMBER: 60/198587
PRIOR FILING DATE: 2000-04-19
PRIOR APPLICATION NUMBER: 60/199614
PRIOR FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: 60/206330
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/206368
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/209832

PRIOR FILING DATE: 2000-06-05
PRIOR APPLICATION NUMBER: 60/218371
PRIOR FILING DATE: 2000-07-13
PRIOR APPLICATION NUMBER: 60/222695
PRIOR FILING DATE: 2000-08-02
PRIOR APPLICATION NUMBER: 60/229896
PRIOR FILING DATE: 2000-09-01
PRIOR APPLICATION NUMBER: 60/230621
PRIOR FILING DATE: 2000-09-05
PRIOR APPLICATION NUMBER: 60/232887
PRIOR FILING DATE: 2000-09-15
PRIOR APPLICATION NUMBER: 60/235147
PRIOR FILING DATE: 2000-09-22
PRIOR APPLICATION NUMBER: 60/261878
PRIOR FILING DATE: 2001-01-12
PRIOR APPLICATION NUMBER: 60/261910
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: 60/261939
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: 60/262150
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: 60/264395
PRIOR FILING DATE: 2001-01-25
PRIOR APPLICATION NUMBER: 60/266421
PRIOR FILING DATE: 2001-02-02
PRIOR APPLICATION NUMBER: 60/267623
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/274399
PRIOR FILING DATE: 2001-03-09
PRIOR APPLICATION NUMBER: 60/280982
PRIOR FILING DATE: 2001-04-03
PRIOR APPLICATION NUMBER: 60/282129
PRIOR FILING DATE: 2001-04-04
PRIOR APPLICATION NUMBER: 60/282199
PRIOR FILING DATE: 2001-04-04
PRIOR APPLICATION NUMBER: 60/290589
PRIOR FILING DATE: 2001-05-09
PRIOR APPLICATION NUMBER: 60/180997
PRIOR FILING DATE: 1998-11-19
PRIOR APPLICATION NUMBER: 60/267213
PRIOR FILING DATE: 1999-03-12
PRIOR APPLICATION NUMBER: 60/380137
PRIOR FILING DATE: 1999-08-25
PRIOR APPLICATION NUMBER: 60/380138
PRIOR FILING DATE: 1999-08-25
PRIOR APPLICATION NUMBER: 60/403297
PRIOR FILING DATE: 1999-10-18
PRIOR APPLICATION NUMBER: 60/423741
PRIOR FILING DATE: 1999-11-10
PRIOR APPLICATION NUMBER: 60/709238
PRIOR FILING DATE: 2000-11-08
PRIOR APPLICATION NUMBER: 60/802706
PRIOR FILING DATE: 2001-03-09
PRIOR APPLICATION NUMBER: 60/872035
PRIOR FILING DATE: 2001-06-01
PRIOR APPLICATION NUMBER: 60/918585
PRIOR FILING DATE: 2001-07-30
PRIOR APPLICATION NUMBER: 60/924419
PRIOR FILING DATE: 2001-08-06
PRIOR APPLICATION NUMBER: 60/927796
PRIOR FILING DATE: 2001-08-09
PRIOR APPLICATION NUMBER: 60/929404
PRIOR FILING DATE: 2001-08-13
PRIOR APPLICATION NUMBER: 60/931836
PRIOR FILING DATE: 2001-08-16
PRIOR APPLICATION NUMBER: 60/941992
PRIOR FILING DATE: 2001-08-28
PRIOR APPLICATION NUMBER: 60/946374
PRIOR FILING DATE: 2001-09-04
PRIOR APPLICATION NUMBER: 60/001054
PRIOR FILING DATE: 2001-11-30
PRIOR APPLICATION NUMBER: 60/052586
PRIOR FILING DATE: 2002-01-15

PRIOR APPLICATION NUMBER: 10/081056
PRIOR FILING DATE: 2002-02-20
PRIOR APPLICATION NUMBER: 10/119480
PRIOR FILING DATE: 2002-04-09

Query Match 6.5%; Score 108, DB 9, Length 877;
Best Local Similarity 24.4%; Pred. No. 2.6;
Matches 64; Conservative 28; Mismatches 94; Indels 76; Gaps 16;

QY 23 GVITTLIAGASKEPTYNSTSGSHRTSSG-----GLIAGSVITDSGVNRY 73
DB 446 GADPICVAGRLS-PGCDGIISSG-RRPDGGCGGDDSTCRIVSGN--LTD-RGGPLGY 500
QY 74 Q---VKQDPTVSKIAORYGLMREIGHINNINSSY-----TITYGWL----- 113
DB 501 QKIMIRAGALRIQIADL-----RPSSNYLALRGPGRSITINGMAVDPPGSGY 548
QY 114 ----TWSGDLKVERISISSGVNT-AHTSPFVAV-----QSSRPVVOQHPAVOKPTPVV 164
DB 549 RAGGTVFRRYRPPREBEKGESLSAEGPTTQPDVYMI FQENPGVFYQYVISPPP---I 605
QY 165 VKKPTPPPVVQ-OP--APVAPPTAPFRATGSGGVMQF-----RYPVQATN 208
DB 606 LENTPEPPVPVQLPETLRVEPPPLAPAPRAPRTGTLQROYRIPQMPAPPHPTPLGSPA 665
QY 209 FVVRPFGTAVGSTVTSNGMW 230
DB 666 AYMKRVGHSACSAS--CGKGYW 685

RESULT 28

US-10-238-183-90
Sequence 90. Application US/10238183
Publication No. US20030073189A1

GENERAL INFORMATION:

APPLICANT: Baker, Kevin
APPLICANT: Baton, Dan
APPLICANT: Filvaroff, Ellen
APPLICANT: Goddard, Audrey
APPLICANT: Grimaldi, J. Christopher
APPLICANT: Gunney, Austin
APPLICANT: Smith, Victoria
APPLICANT: Stephan, Jean-Phillippe
APPLICANT: Matambe, Colin
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin

APPLICANT: Fong, Sherman
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3630R1C11

CURRENT APPLICATION NUMBER: US/10/238, 183

PRIOR FILING DATE: 2002-09-09
PRIOR APPLICATION NUMBER: 10/197942
PRIOR FILING DATE: 2002-07-18
PRIOR APPLICATION NUMBER: 60/059114
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/063046
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/065027
PRIOR FILING DATE: 1997-11-10
PRIOR APPLICATION NUMBER: 60/079689
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/086478
PRIOR FILING DATE: 1998-05-22
PRIOR APPLICATION NUMBER: 60/087607
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/089801
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/090557
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090689
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/091358

PRIOR FILING DATE: 1998-07-01
PRIOR APPLICATION NUMBER: 60/091978
PRIOR FILING DATE: 1998-07-07
PRIOR APPLICATION NUMBER: 60/099803
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/106932
PRIOR FILING DATE: 1998-11-03
PRIOR APPLICATION NUMBER: 60/115554
PRIOR FILING DATE: 1999-01-12
PRIOR APPLICATION NUMBER: 60/119342
PRIOR FILING DATE: 1999-02-09
PRIOR APPLICATION NUMBER: 60/123957
PRIOR FILING DATE: 1999-03-12
PRIOR APPLICATION NUMBER: 60/123972
PRIOR FILING DATE: 1999-03-11
PRIOR APPLICATION NUMBER: 60/127372
PRIOR FILING DATE: 1999-04-01
PRIOR APPLICATION NUMBER: 60/131271
PRIOR FILING DATE: 1999-04-27
PRIOR APPLICATION NUMBER: 60/133459
PRIOR FILING DATE: 1999-05-11
PRIOR APPLICATION NUMBER: 60/135725
PRIOR FILING DATE: 1999-05-25
PRIOR APPLICATION NUMBER: 60/135729
PRIOR FILING DATE: 1999-05-25
PRIOR APPLICATION NUMBER: 60/135750
PRIOR FILING DATE: 1999-05-25
PRIOR APPLICATION NUMBER: 60/138385
PRIOR FILING DATE: 1999-06-09
PRIOR APPLICATION NUMBER: 60/140653
PRIOR FILING DATE: 1999-06-22
PRIOR APPLICATION NUMBER: 60/141037
PRIOR FILING DATE: 1999-06-23
PRIOR APPLICATION NUMBER: 60/144732
PRIOR FILING DATE: 1999-07-20
PRIOR APPLICATION NUMBER: 60/144758
PRIOR FILING DATE: 1999-07-20
PRIOR APPLICATION NUMBER: 60/144790
PRIOR FILING DATE: 1999-07-20
PRIOR APPLICATION NUMBER: 60/145228
PRIOR FILING DATE: 1999-07-20
PRIOR APPLICATION NUMBER: 60/145698
PRIOR FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: 60/146222
PRIOR FILING DATE: 1999-07-28
PRIOR APPLICATION NUMBER: 60/146843
PRIOR FILING DATE: 1999-08-03
PRIOR APPLICATION NUMBER: 60/148188
PRIOR FILING DATE: 1999-08-10
PRIOR APPLICATION NUMBER: 60/148513
PRIOR FILING DATE: 1999-08-12
PRIOR APPLICATION NUMBER: 60/149327
PRIOR FILING DATE: 1999-08-17
PRIOR APPLICATION NUMBER: 60/149395
PRIOR FILING DATE: 1999-08-17
PRIOR APPLICATION NUMBER: 60/150114
PRIOR FILING DATE: 1999-08-20
PRIOR APPLICATION NUMBER: 60/151700
PRIOR FILING DATE: 1999-08-31
PRIOR APPLICATION NUMBER: 60/151724
PRIOR FILING DATE: 1999-08-31
PRIOR APPLICATION NUMBER: 60/162506
PRIOR FILING DATE: 1999-10-29
PRIOR APPLICATION NUMBER: 60/170262
PRIOR FILING DATE: 1999-12-09
PRIOR APPLICATION NUMBER: 60/177118
PRIOR FILING DATE: 2000-01-20
PRIOR APPLICATION NUMBER: 60/179851
PRIOR FILING DATE: 2000-02-02
PRIOR APPLICATION NUMBER: 60/180921
PRIOR FILING DATE: 2000-02-08
PRIOR APPLICATION NUMBER: 60/187202
PRIOR FILING DATE: 2000-03-03

PRIOR APPLICATION NUMBER: 60/198587
PRIOR FILING DATE: 2000-04-18
PRIOR APPLICATION NUMBER: 60/199614
PRIOR FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: 60/206310
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/206368
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/209832
PRIOR FILING DATE: 2000-06-05
PRIOR APPLICATION NUMBER: 60/218371
PRIOR FILING DATE: 2000-07-13
PRIOR APPLICATION NUMBER: 60/222695
PRIOR FILING DATE: 2000-08-02
PRIOR APPLICATION NUMBER: 60/229896
PRIOR FILING DATE: 2000-09-01
PRIOR APPLICATION NUMBER: 60/230621
PRIOR FILING DATE: 2000-09-05
PRIOR APPLICATION NUMBER: 60/232887
PRIOR FILING DATE: 2000-09-15
PRIOR APPLICATION NUMBER: 60/235147
PRIOR FILING DATE: 2000-09-22
PRIOR APPLICATION NUMBER: 60/261878
PRIOR FILING DATE: 2001-01-12
PRIOR APPLICATION NUMBER: 60/261910
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: 60/261939
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: 60/262150
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: 60/264395
PRIOR FILING DATE: 2001-01-25
PRIOR APPLICATION NUMBER: 60/266421
PRIOR FILING DATE: 2001-02-02
PRIOR APPLICATION NUMBER: 60/267623
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/274399
PRIOR FILING DATE: 2001-03-09
PRIOR APPLICATION NUMBER: 60/280982
PRIOR FILING DATE: 2001-04-03
PRIOR APPLICATION NUMBER: 60/282129
PRIOR FILING DATE: 2001-04-04
PRIOR APPLICATION NUMBER: 60/282199
PRIOR FILING DATE: 2001-04-04
PRIOR APPLICATION NUMBER: 60/290589
PRIOR FILING DATE: 2001-05-09
PRIOR APPLICATION NUMBER: 09/180997
PRIOR FILING DATE: 1998-11-19
PRIOR APPLICATION NUMBER: 09/267213
PRIOR FILING DATE: 1999-03-12
PRIOR APPLICATION NUMBER: 09/380137
PRIOR FILING DATE: 1999-08-25
PRIOR APPLICATION NUMBER: 09/380138
PRIOR FILING DATE: 1999-08-25
PRIOR APPLICATION NUMBER: 09/403297
PRIOR FILING DATE: 1999-10-18
PRIOR APPLICATION NUMBER: 09/423741
PRIOR FILING DATE: 1999-11-10
PRIOR APPLICATION NUMBER: 09/709238
PRIOR FILING DATE: 2000-11-08
PRIOR APPLICATION NUMBER: 09/802706
PRIOR FILING DATE: 2001-03-09
PRIOR APPLICATION NUMBER: 09/872035
PRIOR FILING DATE: 2001-06-01
PRIOR APPLICATION NUMBER: 09/918585
PRIOR FILING DATE: 2001-07-30
PRIOR APPLICATION NUMBER: 09/924419
PRIOR FILING DATE: 2001-08-06
PRIOR APPLICATION NUMBER: 09/927796
PRIOR FILING DATE: 2001-08-09
PRIOR APPLICATION NUMBER: 09/929404
PRIOR FILING DATE: 2001-08-13
PRIOR APPLICATION NUMBER: 09/931836

PRIOR FILING DATE: 2001-08-16
PRIOR APPLICATION NUMBER: 09/941992
PRIOR FILING DATE: 2001-08-28
PRIOR APPLICATION NUMBER: 09/946374
PRIOR FILING DATE: 2001-09-04
PRIOR APPLICATION NUMBER: 10/001054
PRIOR FILING DATE: 2001-11-30
PRIOR APPLICATION NUMBER: 10/052586
PRIOR FILING DATE: 2002-01-15
PRIOR APPLICATION NUMBER: 10/081056
PRIOR FILING DATE: 2002-02-20
PRIOR APPLICATION NUMBER: 10/119480
PRIOR FILING DATE: 2002-04-09

Query Match 6.5%; Score 108; DB 9; Length 877;
Best Local Similarity 24.4%; Pred. No. 2.6;
Matches 64; Conservative 28; Mismatches 94; Indels 76; Gaps 16;

QY 23 GVITTCILAGCASKPTYNSTSGSHRTSGSG-----GLAISQVITDSGVNRY 73
DB 446 GAPDVCVAGRLS-PGCDGILGSG-RRPDGGVCGGDDSTCRIVSGN-LTD-RGGLGY 500
QY 74 Q-----VKQGDVSKIAORYGIMREIGHINNLSY-----TIYTGQWL----- 113
DB 501 QKIMIPAGALRLQIAOL-----RPSGNVIALRGPGGRSIIINGMAVDPPGSY 548
QY 114 -----TLMSGDLKVERISISGCVNT-AHTSPVAV-----QSSRPVQOHPAVOKTTPPVV 164
DB 549 RAGGVTFEYVNPPEEGKESLSAEGPTTQVDVYMFQENPNQVFQYVYISSPP---I 605
QY 165 VKKPTPPPVVQ-QP-APVAPVTEAPFATGSGVWQF-----RYPVATN 208
DB 606 LENTPEPPVQQLPELIRVPEPLAPAPRRARTGTILQRVRIQWMPAPPHRTPLGSPA 665
QY 209 PVRRPGTAVAGSTVNSGMW 230
DB 666 AYKRVGHSACSAS--CGKGVW 685

RESULT 29
US-10-238-283-90
Sequence 90, Application US/10238283
Publication No. US20030073190A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin
APPLICANT: Filvaroff, Ellen
APPLICANT: Goddard, Audrey
APPLICANT: Grimaldi, J. Christopher
APPLICANT: Gurney, Austin
APPLICANT: Smith, Victoria
APPLICANT: Stephan, Jean-Philippe
APPLICANT: Wacande, Colin
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin
APPLICANT: Fong, Sherman
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3630R1C15
CURRENT APPLICATION NUMBER: US/10/238,283
PRIOR APPLICATION NUMBER: 10/197942
PRIOR FILING DATE: 2002-07-18
PRIOR APPLICATION NUMBER: 60/059114
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/063046
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/065027
PRIOR FILING DATE: 1997-11-10
PRIOR APPLICATION NUMBER: 60/079689
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/086478
PRIOR FILING DATE: 1998-05-22

PRIOR APPLICATION NUMBER: 60/087607
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/089801
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/090557
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090689
PRIOR FILING DATE: 1998-06-25
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 116
SEQ ID NO 90
LENGTH: 877
TYPE: PRT
ORGANISM: Homo Sapien
US-10-238-283-90

Query Match 6.5%; Score 108; DB 9; Length 877;
Best Local Similarity 24.4%; Pred. No. 2.6;
Matches 64; Conservative 28; Mismatches 94; Indels 76; Gaps 16;

QY 23 GVTTCTIACGACSKPTYNSTSGSHRTSGSG-----GLAIGSVITDSQGVNRY 73
DB 446 GARDICVAGRCLT-PGCDGILGSG-RRPDGCGVCGDDSTCRVSGN--LTD-RGGLGY 500
QY 74 Q-----VKQDTSKIAQRYGLNREIGHINNLSY-----TYYGQWL----- 113
DB 501 QKILMTFAGALRLQIQL-----RPSNYIALRGPGGRSLINNMWAVDPGGSY 548
QY 114 -----TWSGDLKXERSISSGVNT-AHTSPVAV-----QSRPPVOQHPAVQKTPPVV 164
DB 549 RAGGTVRNRRPREBEKGSLSAEGTTPVDVYMFQENPVFYQYVSSPP--I 605
QY 165 VKRTPPPVVO-QP-APVAPVTEAPFATGSSGVNQF-----RYVGATN 208
DB 606 LENTPEPPVQLOPELIRVEPLAPRPARPTGTLQKQVRIPQMPAPRPTPLGSPA 665
QY 209 PVVRRGTATVAGSTVTSNGM 230
DB 666 AYMKRVGHSACSAS--CGKGVW 685

RESULT 30
US-10-238-370-90
Sequence 90, Application US/10238370
Publication No. US20030073191A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin
APPLICANT: Baton, Dan
APPLICANT: Filvaroff, Ellen
APPLICANT: Goddard, Audrey
APPLICANT: Grimaldi, J. Christopher
APPLICANT: Gurney, Austin
APPLICANT: Smith, Victoria
APPLICANT: Stephan, Jean-Phillippe
APPLICANT: Watanabe, Colin
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin
APPLICANT: Fong, Sherman
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3630R1C10
CURRENT APPLICATION NUMBER: US/10/238, 370
CURRENT FILING DATE: 2002-09-09
PRIOR APPLICATION NUMBER: 10/197942
PRIOR FILING DATE: 2002-07-18
PRIOR APPLICATION NUMBER: 60/059114
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/063046
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/065027
PRIOR FILING DATE: 1997-11-10
PRIOR APPLICATION NUMBER: 60/079689
PRIOR FILING DATE: 1998-03-27

PRIOR APPLICATION NUMBER: 60/086478
PRIOR FILING DATE: 1998-05-22
PRIOR APPLICATION NUMBER: 60/087607
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/089801
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/090557
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090689
PRIOR FILING DATE: 1998-06-25
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 116
SEQ ID NO 90
LENGTH: 877
TYPE: PRT
ORGANISM: Homo Sapien
US-10-238-370-90

Query Match 6.5%; Score 108; DB 9; Length 877;
Best Local Similarity 24.4%; Pred. No. 2.6;
Matches 64; Conservative 28; Mismatches 94; Indels 76; Gaps 16;

QY 23 GVTTCTIACGACSKPTYNSTSGSHRTSGSG-----GLAIGSVITDSQGVNRY 73
DB 446 GARDICVAGRCLT-PGCDGILGSG-RRPDGCGVCGDDSTCRVSGN--LTD-RGGLGY 500
QY 74 Q-----VKQDTSKIAQRYGLNREIGHINNLSY-----TYYGQWL----- 113
DB 501 QKILMTFAGALRLQIQL-----RPSNYIALRGPGGRSLINNMWAVDPGGSY 548
QY 114 -----TWSGDLKXERSISSGVNT-AHTSPVAV-----QSRPPVOQHPAVQKTPPVV 164
DB 549 RAGGTVRNRRPREBEKGSLSAEGTTPVDVYMFQENPVFYQYVSSPP--I 605
QY 165 VKRTPPPVVO-QP-APVAPVTEAPFATGSSGVNQF-----RYVGATN 208
DB 606 LENTPEPPVQLOPELIRVEPLAPRPARPTGTLQKQVRIPQMPAPRPTPLGSPA 665
QY 209 PVVRRGTATVAGSTVTSNGM 230
DB 666 AYMKRVGHSACSAS--CGKGVW 685

RESULT 31
US-10-245-055-90
Sequence 90, Application US/10245055
Publication No. US20030073192A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin
APPLICANT: Baton, Dan
APPLICANT: Filvaroff, Ellen
APPLICANT: Goddard, Audrey
APPLICANT: Grimaldi, J. Christopher
APPLICANT: Gurney, Austin
APPLICANT: Smith, Victoria
APPLICANT: Stephan, Jean-Phillippe
APPLICANT: Watanabe, Colin
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin
APPLICANT: Fong, Sherman
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3630R1C88
CURRENT APPLICATION NUMBER: US/10/245, 055
CURRENT FILING DATE: 2002-09-16
PRIOR APPLICATION NUMBER: 10/197942
PRIOR FILING DATE: 2002-07-18
PRIOR APPLICATION NUMBER: 60/059114
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/063046
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/065027
PRIOR FILING DATE: 1997-11-10

PRIOR APPLICATION NUMBER: 60/079689
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/086478
PRIOR FILING DATE: 1998-05-22
PRIOR APPLICATION NUMBER: 60/087607
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/089801
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/090557
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090689
PRIOR FILING DATE: 1998-06-25
Remaining prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 116
SEQ ID NO 90
LENGTH: 877
TYPE: PRT
ORGANISM: Homo Sapien
US-10-245-055-90

Query Match 6.5%; Score 108; DB 9; Length 877;
Best Local Similarity 24.4%; Pred. No. 2.6;
Matches 64; Conservative 28; Mismatches 94; Indels 76; Gaps 16;

QY 23 GVITTCILAGCASKPTVNSTSGSHRTSGSG-----GLAISQVITTSQGVPNRY 73
DB 446 GADPICVAGRCILS-PCGCGILGSG-RRPDGCGVCGDDSTCRVSGN-LTD-RGGPLGY 500
QY 74 Q---VKQGDVSKIAQRYGLNWRIGHINNLSY-----TIYGOYL----- 113
DB 501 QKILMIPAGALRLQIALQ-----RPSNVLALRGPGGRSLINGNNAVDPGSGY 548
QY 114 ----TLWSGDLKVEREISISGVNT-AHTSPVAV-----QSSRPVQGHAPVQKPTPPVV 164
DB 549 RAGGTIVRYNRPPEBEKGESLSAEGPTQPDVYMIPEENPGVYQYVYISSPP--I 605
QY 165 VKKTPPPVVO-QP-APVAPVTEAPPATSSGVNQF-----RYVGATN 208
DB 606 LENTPEPPVPOQLPELIRVEPLAPRPARPTGTLQROVRIPOMPAPPHRTPLGSPA 665
QY 209 PVVRRFGTAVAGSTVTSNGMW 230
DB 666 AYWKRVGHSACSAS--CGKGVW 685

RESULT 32

US-10-245-147-90
Sequence 90, Application US/10245147
Publication No. US20030073193A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin
APPLICANT: Baton, Dan
APPLICANT: Filvaroff, Ellen
APPLICANT: Goddard, Audrey
APPLICANT: Grimaldi, J. Christopher
APPLICANT: Guiney, Austin
APPLICANT: Smith, Victoria
APPLICANT: Stephan, Jean-Philippe
APPLICANT: Watanabe, Collin
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin
APPLICANT: Fong, Sherman
TITLE OF INVENTION: ACIDS ENCODING THE SAME
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3630R1C72
CURRENT APPLICATION NUMBER: US/10/245,147
CURRENT FILING DATE: 2002-09-16
PRIOR APPLICATION NUMBER: 10/197942
PRIOR FILING DATE: 2002-07-18
PRIOR APPLICATION NUMBER: 60/059114
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/063046
PRIOR FILING DATE: 1997-10-24

PRIOR APPLICATION NUMBER: 60/065027
PRIOR FILING DATE: 1997-11-10
PRIOR APPLICATION NUMBER: 60/079689
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/086478
PRIOR FILING DATE: 1998-05-22
PRIOR APPLICATION NUMBER: 60/087607
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/089801
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/090557
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090689
PRIOR FILING DATE: 1998-06-25
Remaining prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 116
SEQ ID NO 90
LENGTH: 877
TYPE: PRT
ORGANISM: Homo Sapien
US-10-245-147-90

Query Match 6.5%; Score 108; DB 9; Length 877;
Best Local Similarity 24.4%; Pred. No. 2.6;
Matches 64; Conservative 28; Mismatches 94; Indels 76; Gaps 16;

QY 23 GVITTCILAGCASKPTVNSTSGSHRTSGSG-----GLAISQVITTSQGVPNRY 73
DB 446 GADPICVAGRCILS-PCGCGILGSG-RRPDGCGVCGDDSTCRVSGN-LTD-RGGPLGY 500
QY 74 Q---VKQGDVSKIAQRYGLNWRIGHINNLSY-----TIYGOYL----- 113
DB 501 QKILMIPAGALRLQIALQ-----RPSNVLALRGPGGRSLINGNNAVDPGSGY 548
QY 114 ----TLWSGDLKVEREISISGVNT-AHTSPVAV-----QSSRPVQGHAPVQKPTPPVV 164
DB 549 RAGGTIVRYNRPPEBEKGESLSAEGPTQPDVYMIPEENPGVYQYVYISSPP--I 605
QY 165 VKKTPPPVVO-QP-APVAPVTEAPPATSSGVNQF-----RYVGATN 208
DB 606 LENTPEPPVPOQLPELIRVEPLAPRPARPTGTLQROVRIPOMPAPPHRTPLGSPA 665
QY 209 PVVRRFGTAVAGSTVTSNGMW 230
DB 666 AYWKRVGHSACSAS--CGKGVW 685

RESULT 33

US-10-245-730-90
Sequence 90, Application US/10245730
Publication No. US20030073194A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin
APPLICANT: Baton, Dan
APPLICANT: Filvaroff, Ellen
APPLICANT: Goddard, Audrey
APPLICANT: Grimaldi, J. Christopher
APPLICANT: Guiney, Austin
APPLICANT: Smith, Victoria
APPLICANT: Stephan, Jean-Philippe
APPLICANT: Watanabe, Collin
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin
APPLICANT: Fong, Sherman
TITLE OF INVENTION: ACIDS ENCODING THE SAME
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3630R1C85
CURRENT APPLICATION NUMBER: US/10/245,730
CURRENT FILING DATE: 2002-09-16
PRIOR APPLICATION NUMBER: 10/197942
PRIOR FILING DATE: 2002-07-18
PRIOR APPLICATION NUMBER: 60/059114
PRIOR FILING DATE: 1997-09-17

;; PRIOR APPLICATION NUMBER: 60/063046
;; PRIOR FILING DATE: 1997-10-24
;; PRIOR APPLICATION NUMBER: 60/065027
;; PRIOR FILING DATE: 1997-11-10
;; PRIOR APPLICATION NUMBER: 60/079689
;; PRIOR FILING DATE: 1998-03-27
;; PRIOR APPLICATION NUMBER: 60/086478
;; PRIOR FILING DATE: 1998-05-22
;; PRIOR APPLICATION NUMBER: 60/087607
;; PRIOR FILING DATE: 1998-06-02
;; PRIOR APPLICATION NUMBER: 60/089801
;; PRIOR FILING DATE: 1998-06-18
;; PRIOR APPLICATION NUMBER: 60/090557
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090689
;; PRIOR FILING DATE: 1998-06-25
;; Remaining Prior Application data removed - See File Wrapper or PALM.
;; NUMBER OF SEQ ID NOS: 116
;; SEQ ID NO 90
;; LENGTH: 877
;; TYPE: PRT
;; ORGANISM: Homo Sapien
US-10-245-730-90

Query Match 6.5%; Score 108; DB 9; Length 877;
Best Local Similarity 24.4%; Pred. No. 2.6;
Matches 64; Conservative 28; Mismatches 94; Indels 76; Gaps 16;
QY 23 GVITTCILACAKPTYNSTSGSHRTSGSG-----GLAIGQVITDSQGVPNRY 73
DB 446 GARDICVAGRCLS-PGCDGILGSG-RRPDGCGVCGDDSTCRIVSGN--LTD-RGGLG 500
QY 74 Q---VKQGPVSKIAQRYGLNREIGHINNINSSY-----TITYGWL----- 113
DB 501 QKILMTIPAGALRIQIQL-----RPSNYIALRGPGRSITINGMMAVDPGSGY 548
QY 114 ----TLMSGDLKVERSISSGVNT-AHTPSFVAV-----QSSRPVQOHRAVQKPTPPVV 164
DB 549 RAGTVERVYRPPREBEKGSLSAEGPTTOPVDVYMIFOENPGVFQYVYISSPP--I 605
QY 165 VKKPTPPPVQ-QP--APVAPVTEAPFATGSGVQF-----RYVQATN 208
DB 606 LENTPEPPVPQQLPELIRVPEPLAPRAPRARTGTILQRYVRIPOMAPRPHRTPLGSPA 665
QY 209 PVVRPFGATVAGSTVTSNGMW 230
DB 666 AYMKRVGSHSACSAS--CGKGW 685

RESULT 34

US-10-245-739-90
;; Sequence 90, Application US/10245739
;; Publication No. US20030073195A1
;; GENERAL INFORMATION:
;; APPLICANT: Baker, Kevin
;; APPLICANT: Filvaroff, Ellen
;; APPLICANT: Goddard, Audrey
;; APPLICANT: Grimaldi, J. Christopher
;; APPLICANT: Gurney, Austin
;; APPLICANT: Smith, Victoria
;; APPLICANT: Stephan, Jean-Phillippe
;; APPLICANT: Watanabe, Colin
;; APPLICANT: Wood, William
;; APPLICANT: Zhang, Zemin
;; APPLICANT: Fong, Sherman
;; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;; FILE REFERENCE: P363081C86
;; CURRENT FILING DATE: US/10/245, 739
;; CURRENT APPLICATION NUMBER: 2002-09-16
;; PRIOR APPLICATION NUMBER: 10/197942
;; PRIOR FILING DATE: 2002-07-18

;; PRIOR APPLICATION NUMBER: 60/059114
;; PRIOR FILING DATE: 1997-09-17
;; PRIOR APPLICATION NUMBER: 60/063046
;; PRIOR FILING DATE: 1997-10-24
;; PRIOR APPLICATION NUMBER: 60/065027
;; PRIOR FILING DATE: 1997-11-10
;; PRIOR APPLICATION NUMBER: 60/079689
;; PRIOR FILING DATE: 1998-03-27
;; PRIOR APPLICATION NUMBER: 60/086478
;; PRIOR FILING DATE: 1998-05-22
;; PRIOR APPLICATION NUMBER: 60/087607
;; PRIOR FILING DATE: 1998-06-02
;; PRIOR APPLICATION NUMBER: 60/089801
;; PRIOR FILING DATE: 1998-06-18
;; PRIOR APPLICATION NUMBER: 60/090557
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090689
;; PRIOR FILING DATE: 1998-06-25
;; Remaining Prior Application data removed - See File Wrapper or PALM.
;; NUMBER OF SEQ ID NOS: 116
;; SEQ ID NO 90
;; LENGTH: 877
;; TYPE: PRT
;; ORGANISM: Homo Sapien
US-10-245-739-90

Query Match 6.5%; Score 108; DB 9; Length 877;
Best Local Similarity 24.4%; Pred. No. 2.6;
Matches 64; Conservative 28; Mismatches 94; Indels 76; Gaps 16;
QY 23 GVITTCILACAKPTYNSTSGSHRTSGSG-----GLAIGQVITDSQGVPNRY 73
DB 446 GARDICVAGRCLS-PGCDGILGSG-RRPDGCGVCGDDSTCRIVSGN--LTD-RGGLG 500
QY 74 Q---VKQGPVSKIAQRYGLNREIGHINNINSSY-----TITYGWL----- 113
DB 501 QKILMTIPAGALRIQIQL-----RPSNYIALRGPGRSITINGMMAVDPGSGY 548
QY 114 ----TLMSGDLKVERSISSGVNT-AHTPSFVAV-----QSSRPVQOHRAVQKPTPPVV 164
DB 549 RAGTVERVYRPPREBEKGSLSAEGPTTOPVDVYMIFOENPGVFQYVYISSPP--I 605
QY 165 VKKPTPPPVQ-QP--APVAPVTEAPFATGSGVQF-----RYVQATN 208
DB 606 LENTPEPPVPQQLPELIRVPEPLAPRAPRARTGTILQRYVRIPOMAPRPHRTPLGSPA 665
QY 209 PVVRPFGATVAGSTVTSNGMW 230
DB 666 AYMKRVGSHSACSAS--CGKGW 685

RESULT 35

US-10-246-210-90
;; Sequence 90, Application US/10246210
;; Publication No. US20030073196A1
;; GENERAL INFORMATION:
;; APPLICANT: Baker, Kevin
;; APPLICANT: Filvaroff, Ellen
;; APPLICANT: Goddard, Audrey
;; APPLICANT: Grimaldi, J. Christopher
;; APPLICANT: Gurney, Austin
;; APPLICANT: Smith, Victoria
;; APPLICANT: Stephan, Jean-Phillippe
;; APPLICANT: Watanabe, Colin
;; APPLICANT: Wood, William
;; APPLICANT: Zhang, Zemin
;; APPLICANT: Fong, Sherman
;; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;; FILE REFERENCE: P363081C121
;; CURRENT FILING DATE: US/10/246, 210
;; CURRENT APPLICATION NUMBER: 2002-09-18

```

; PRIOR APPLICATION NUMBER: 10/197942
; PRIOR FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/059114
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/063046
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/065027
; PRIOR FILING DATE: 1997-11-10
; PRIOR APPLICATION NUMBER: 60/079689
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/086478
; PRIOR FILING DATE: 1998-05-22
; PRIOR APPLICATION NUMBER: 60/087607
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/089801
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090689
; PRIOR FILING DATE: 1998-06-25
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 116
; SEQ ID NO 90
; LENGTH: 877
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-246-210-90
```

Query Match 6.5%; Score 108; DB 9; Length 877;
Best Local Similarity 24.4%; Pred. No. 2.6;

Matches 64; Conservative 28; Mismatches 94; Indels 76; Gaps 16;

```

QY 23 GVITTCILAGCASKPTYNSTSGSHRTSGSG-----GLAISQVITDSQGVPNRY 73
DB 446 GADPICVAGRCLS-FGDDGILGSG-RRPDGCGVCGGDDSTCRIVSGN--LTD-RGGLIY 500
QY 74 Q-----VKQDPTVSKIAQRYGLNREIGHINNLSY-----TIYQWL----- 113
DB 501 QKILMIPAGLRILQIALQ-----RPSNLTALRGPGGRSIIINGNAVDPGSGY 548
QY 114 ----TLMSGDLKVERSISSGVNT-AHTSPYAV-----OSSRPVQCHPAVQKPPPPV 164
DB 549 RAGTGYFRYRPPRECKGESLSAEGPTTPVDVYMIFOENRGVYGVYSSPP--I 605
QY 165 VKKPTPTPPVYQ-QP--APVAPVTEAPPATGSSGVMP-----RYVGATN 208
DB 606 LENTPEPPVPQLOPEILRVEPLAPAPRAPRTPTGLQOVRIPOMPAPPHRTPLGSPA 665
QY 209 PTVRRFGTATVAGSTVTSNGMW 230
DB 666 AYWKRVGHSACSAS--CGKGVW 685
```

RESULT 36

US-10-239-196-90

; Sequence 90, Application US/10239196

; Publication No. US20030074735A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin

; APPLICANT: Baton, Den

; APPLICANT: Filvaroff, Ellen

; APPLICANT: Goddard, Audrey

; APPLICANT: Grimaldi, J. Christopher

; APPLICANT: Guirey, Austin

; APPLICANT: Smith, Victoria

; APPLICANT: Stephan, Jean-Phillippe

; APPLICANT: Watande, Collin

; APPLICANT: Wood, William

; APPLICANT: Zhang, Zemin

; APPLICANT: Fong, Sherman

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

; FILE REFERENCE: P3630R1C8

```

; CURRENT APPLICATION NUMBER: US/10/239,196
; PRIOR FILING DATE: 2002-09-09
; PRIOR APPLICATION NUMBER: 10/197942
; PRIOR FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/059114
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/063046
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/065027
; PRIOR FILING DATE: 1997-11-10
; PRIOR APPLICATION NUMBER: 60/079689
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/086478
; PRIOR FILING DATE: 1998-05-22
; PRIOR APPLICATION NUMBER: 60/087607
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/089801
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090689
; PRIOR FILING DATE: 1998-06-25
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 116
; SEQ ID NO 90
; LENGTH: 877
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-239-196-90
```

Query Match 6.5%; Score 108; DB 9; Length 877;
Best Local Similarity 24.4%; Pred. No. 2.6;

Matches 64; Conservative 28; Mismatches 94; Indels 76; Gaps 16;

```

QY 23 GVITTCILAGCASKPTYNSTSGSHRTSGSG-----GLAISQVITDSQGVPNRY 73
DB 446 GADPICVAGRCLS-FGDDGILGSG-RRPDGCGVCGGDDSTCRIVSGN--LTD-RGGLIY 500
QY 74 Q-----VKQDPTVSKIAQRYGLNREIGHINNLSY-----TIYQWL----- 113
DB 501 QKILMIPAGLRILQIALQ-----RPSNLTALRGPGGRSIIINGNAVDPGSGY 548
QY 114 ----TLMSGDLKVERSISSGVNT-AHTSPYAV-----OSSRPVQCHPAVQKPPPPV 164
DB 549 RAGTGYFRYRPPRECKGESLSAEGPTTPVDVYMIFOENRGVYGVYSSPP--I 605
QY 165 VKKPTPTPPVYQ-QP--APVAPVTEAPPATGSSGVMP-----RYVGATN 208
DB 606 LENTPEPPVPQLOPEILRVEPLAPAPRAPRTPTGLQOVRIPOMPAPPHRTPLGSPA 665
QY 209 PTVRRFGTATVAGSTVTSNGMW 230
DB 666 AYWKRVGHSACSAS--CGKGVW 685
```

RESULT 37

US-10-243-024-90

; Sequence 90, Application US/10243024

; Publication No. US20030077741A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin

; APPLICANT: Baton, Den

; APPLICANT: Filvaroff, Ellen

; APPLICANT: Goddard, Audrey

; APPLICANT: Grimaldi, J. Christopher

; APPLICANT: Guirey, Austin

; APPLICANT: Smith, Victoria

; APPLICANT: Stephan, Jean-Phillippe

; APPLICANT: Watande, Collin

; APPLICANT: Wood, William

; APPLICANT: Zhang, Zemin

; APPLICANT: Fong, Sherman

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

; FILE REFERENCE: P3630R1C8

```

; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3630RIC38
; CURRENT APPLICATION NUMBER: US/10/243,024
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: 10/197942
; PRIOR FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/059114
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/063046
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/065027
; PRIOR FILING DATE: 1997-11-10
; PRIOR APPLICATION NUMBER: 60/079689
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/086478
; PRIOR FILING DATE: 1998-05-22
; PRIOR APPLICATION NUMBER: 60/087607
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/089801
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090689
; PRIOR FILING DATE: 1998-06-25
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 116
; SEQ ID NO 90
; LENGTH: 877
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-243-024-90

```

Query Match Best Local Similarity 6.5%; Score 108; DB 9; Length 877;

Matches 64; Conservative 28; Mismatches 94; Indels 76; Gaps 16;

```

QY 23 GVITTCILACGACAPRTYNTSGSGSHRTSGG-----GLAIGSQVITDSQGVNRY 73
DB 446 GADPICVAGRLS-PCGDDGLISG-RRPDGCGVCGGDDSTCRIVSGN--LTD-RGPGLY 500
QY 74 Q-----VKQGDVSKIAQRYGLNMRREIGHINNLSY-----TYTGQWL----- 113
DB 501 QKILMIPAGLRILQIQL-----RPSNNTALRGPGGRSLINGMWAVDPPGSY 548
QY 114 -----TMSGDLKVERNSISSGVT-AHTSPVAV-----QSSRPVQGHNAVQKTPPVV 164
DB 549 RAGGTVERYNRPREEKGESLSAEGPTTPVDVYMIFOENPGVFQYVYSSPP---I 605
QY 165 VKKPTPPPVVQ-OP--APVAPVTEAPFATGSSGVNQF-----RYPVGATN 208
DB 606 LENTPEPPVQLOPELIRVEPLAPAPRPARPTGTLQROVRIQWAPAPPHRTPLSPA 665
QY 209 PVVRRFGTATVAGSTVTSNGM 230
DB 666 AYWKRVGHSACSAS--CGKGW 685

```

RESULT 38
US-10-243-409-90

```

; Sequence 90, Application US/10243409
; Publication No. US2003007742A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin
; APPLICANT: Eaton, Dan
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Goddard, Audrey
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin
; APPLICANT: Smith, Victoria
; APPLICANT: Stephan, Jean-Phillippe
; APPLICANT: Watanabe, Colin
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin

```

```

; APPLICANT: Fong, Sherman
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3630RIC41
; CURRENT APPLICATION NUMBER: US/10/243,409
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: 10/197942
; PRIOR FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/059114
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/063046
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/065027
; PRIOR FILING DATE: 1997-11-10
; PRIOR APPLICATION NUMBER: 60/079689
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/086478
; PRIOR FILING DATE: 1998-05-22
; PRIOR APPLICATION NUMBER: 60/087607
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/089801
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090689
; PRIOR FILING DATE: 1998-06-25
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 116
; SEQ ID NO 90
; LENGTH: 877
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-243-409-90

```

Query Match Best Local Similarity 6.5%; Score 108; DB 9; Length 877;

Matches 64; Conservative 28; Mismatches 94; Indels 76; Gaps 16;

```

QY 23 GVITTCILACGACAPRTYNTSGSGSHRTSGG-----GLAIGSQVITDSQGVNRY 73
DB 446 GADPICVAGRLS-PCGDDGLISG-RRPDGCGVCGGDDSTCRIVSGN--LTD-RGPGLY 500
QY 74 Q-----VKQGDVSKIAQRYGLNMRREIGHINNLSY-----TYTGQWL----- 113
DB 501 QKILMIPAGLRILQIQL-----RPSNNTALRGPGGRSLINGMWAVDPPGSY 548
QY 114 -----TMSGDLKVERNSISSGVT-AHTSPVAV-----QSSRPVQGHNAVQKTPPVV 164
DB 549 RAGGTVERYNRPREEKGESLSAEGPTTPVDVYMIFOENPGVFQYVYSSPP---I 605
QY 165 VKKPTPPPVVQ-OP--APVAPVTEAPFATGSSGVNQF-----RYPVGATN 208
DB 606 LENTPEPPVQLOPELIRVEPLAPAPRPARPTGTLQROVRIQWAPAPPHRTPLSPA 665
QY 209 PVVRRFGTATVAGSTVTSNGM 230
DB 666 AYWKRVGHSACSAS--CGKGW 685

```

RESULT 39
US-10-245-033-90

```

; Sequence 90, Application US/10245033
; Publication No. US20030078401A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin
; APPLICANT: Eaton, Dan
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Goddard, Audrey
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin
; APPLICANT: Smith, Victoria
; APPLICANT: Stephan, Jean-Phillippe
; APPLICANT: Watanabe, Colin

```

```
/ APPLICANT: Wood,William
/ APPLICANT: Zhang,Zemin
/ APPLICANT: Fong,Sherman
/ TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
/ FILE REFERENCE: P3630R1C84
/ CURRENT APPLICATION NUMBER: US/10/245,033
/ CURRENT FILING DATE: 2002-09-16
/ PRIOR APPLICATION NUMBER: 10/197942
/ PRIOR FILING DATE: 2002-07-18
/ PRIOR APPLICATION NUMBER: 60/059114
/ PRIOR FILING DATE: 1997-09-17
/ PRIOR APPLICATION NUMBER: 60/063046
/ PRIOR FILING DATE: 1997-10-24
/ PRIOR APPLICATION NUMBER: 60/065027
/ PRIOR FILING DATE: 1997-11-10
/ PRIOR APPLICATION NUMBER: 60/079689
/ PRIOR FILING DATE: 1998-03-27
/ PRIOR APPLICATION NUMBER: 60/086478
/ PRIOR FILING DATE: 1998-05-22
/ PRIOR APPLICATION NUMBER: 60/087607
/ PRIOR FILING DATE: 1998-06-02
/ PRIOR APPLICATION NUMBER: 60/089801
/ PRIOR FILING DATE: 1998-06-18
/ PRIOR APPLICATION NUMBER: 60/090557
/ PRIOR FILING DATE: 1998-06-24
/ PRIOR APPLICATION NUMBER: 60/090689
/ PRIOR FILING DATE: 1998-06-25
/ Remaining Prior Application data removed - See File Wrapper or PALM.
/ NUMBER OF SEQ ID NOS: 116
/ SEQ ID NO 90
/ LENGTH: 877
/ TYPE: PRT
/ ORGANISM: Homo Sapien
US-10-245-033-90

Query Match      6.5%; Score 108; DB 9; Length 877;
Best Local Similarity 24.4%; Pred. No. 2.6;
Matches 64; Conservative 28; Mismatches 94; Indels 76; Gaps 16;

QY 23 GVTTTCILAGCASKPTYNSTSGSGSHRTSGSG-----GLAIGQVITDSQGVNRY 73
DB 446 GAPDVCVAGRCLLS-PCDDGILGSG-RRPDGCGVCGDDSTCRIVSGN--LTD-RGGPLGY 500
QY 74 Q-----VKQGDVYSKIQRYGLNMRREIGHINNLSY-----TIYGOWL----- 113
DB 501 OKLIMIPAGALRLQIQL-----RPSNLYALRGPGGRSIIINGNMAVDPPGSY 548
QY 114 ----TMSGDLKVERISISSGVNT-AHTPSPAV-----OSSRPVQCHRAVQKPTPPVV 164
DB 549 RAGGTFRNRNPREBKESLSABGPTTOPVDVYMIFOENPGVYQVYISSPP---I 605
QY 165 VKKPTPTPPVQ-QP-APVAPVTEAPFATGSSGMQF-----RYPGATN 208
DB 606 LENPTPEPPVQLOPELIRVEPPLAPRPARPTGTLQROVRLPQMPAPRHPRTPLGSPA 665
QY 209 PVRPRFGTATVAGSTVTSNGMW 230
DB 666 AYMKRVGHSACSAS--CGKGVW 685

RESULT 40
US-10-245-621-90
/ Sequence 90, Application US/10245621
/ Publication No. US20030077743A1
/ GENERAL INFORMATION:
/ APPLICANT: Baker, Kevin
/ APPLICANT: Baton, Dan
/ APPLICANT: Filvaroff, Ellen
/ APPLICANT: Goddard, Audrey
/ APPLICANT: Grimaldi, J. Christopher
/ APPLICANT: Gurney, Austin
/ APPLICANT: Smith, Victoria
```

```
/ APPLICANT: Stephan,Jean-Phillippe
/ APPLICANT: Watande,Colin
/ APPLICANT: Wood,William
/ APPLICANT: Fong,Sherman
/ TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
/ FILE REFERENCE: P3630R1C104
/ CURRENT APPLICATION NUMBER: US/10/245,621
/ CURRENT FILING DATE: 2002-09-17
/ PRIOR APPLICATION NUMBER: 10/197942
/ PRIOR FILING DATE: 2002-07-18
/ PRIOR APPLICATION NUMBER: 60/059114
/ PRIOR FILING DATE: 1997-09-17
/ PRIOR APPLICATION NUMBER: 60/063046
/ PRIOR FILING DATE: 1997-10-24
/ PRIOR APPLICATION NUMBER: 60/065027
/ PRIOR FILING DATE: 1997-11-10
/ PRIOR APPLICATION NUMBER: 60/079689
/ PRIOR FILING DATE: 1998-03-27
/ PRIOR APPLICATION NUMBER: 60/086478
/ PRIOR FILING DATE: 1998-05-22
/ PRIOR APPLICATION NUMBER: 60/087607
/ PRIOR FILING DATE: 1998-06-02
/ PRIOR APPLICATION NUMBER: 60/089801
/ PRIOR FILING DATE: 1998-06-18
/ PRIOR APPLICATION NUMBER: 60/090557
/ PRIOR FILING DATE: 1998-06-24
/ PRIOR APPLICATION NUMBER: 60/090689
/ PRIOR FILING DATE: 1998-06-25
/ Remaining Prior Application data removed - See File Wrapper or PALM.
/ NUMBER OF SEQ ID NOS: 116
/ SEQ ID NO 90
/ LENGTH: 877
/ TYPE: PRT
/ ORGANISM: Homo Sapien
US-10-245-621-90

Query Match      6.5%; Score 108; DB 9; Length 877;
Best Local Similarity 24.4%; Pred. No. 2.6;
Matches 64; Conservative 28; Mismatches 94; Indels 76; Gaps 16;

QY 23 GVTTTCILAGCASKPTYNSTSGSGSHRTSGSG-----GLAIGQVITDSQGVNRY 73
DB 446 GAPDVCVAGRCLLS-PCDDGILGSG-RRPDGCGVCGDDSTCRIVSGN--LTD-RGGPLGY 500
QY 74 Q-----VKQGDVYSKIQRYGLNMRREIGHINNLSY-----TIYGOWL----- 113
DB 501 OKLIMIPAGALRLQIQL-----RPSNLYALRGPGGRSIIINGNMAVDPPGSY 548
QY 114 ----TMSGDLKVERISISSGVNT-AHTPSPAV-----OSSRPVQCHRAVQKPTPPVV 164
DB 549 RAGGTFRNRNPREBKESLSABGPTTOPVDVYMIFOENPGVYQVYISSPP---I 605
QY 165 VKKPTPTPPVQ-QP-APVAPVTEAPFATGSSGMQF-----RYPGATN 208
DB 606 LENPTPEPPVQLOPELIRVEPPLAPRPARPTGTLQROVRLPQMPAPRHPRTPLGSPA 665
QY 209 PVRPRFGTATVAGSTVTSNGMW 230
DB 666 AYMKRVGHSACSAS--CGKGVW 685

Search completed: July 8, 2003, 11:17:11
Job time : 56 secs
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 8, 2003, 11:02:47 ; Search time 41 Seconds
(without alignments)
755.007 Million cell updates/sec

Title: US-10-018-706-2
Perfect score: 1670
Sequence: 1 MVTITAINSONOKPIKRLG.....LFEFRISRGNGVYDPLTVLK 322

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	298	17.8	297	2 S55063	conserved hypochet
2	268	16.0	314	2 T50918	hypothetical prote
3	253	15.1	242	2 A82754	lipoprotein XF0855
4	238	14.3	259	2 A65070	hypothetical prote
5	238	14.3	259	2 B91096	probable lipoprote
6	234	14.0	231	2 B83031	conserved hypochet
7	232.5	13.9	250	2 AD0871	probable lipoprote
8	219.5	13.1	405	2 E4087	lipoprotein D homo
9	219.5	13.1	454	2 E75291	probable cell wall
10	218.5	13.1	333	2 AF0407	lipoprotein (impor
11	210	12.6	373	2 AH0855	lipoprotein N1pD V
12	208.5	12.5	311	2 H82311	lipoprotein N1pD V
13	205	12.3	427	2 A13386	lipoprotein nlpD l
14	203	12.2	415	2 G81864	probable membrane
15	202.5	12.1	379	2 B55522	lipoprotein D prec
16	202.5	12.1	379	2 D91078	probable lipoprote
17	202.5	12.1	379	2 B85923	lipoprotein (impor
18	198	11.9	409	2 H81079	lipoprotein N1pD,
19	196	11.7	615	2 G87496	peptidase, M23/M37
20	188.5	11.3	537	2 AF2785	lipoprotein (impor
21	188.5	11.3	562	2 H97564	43K antigen (Afl57
22	171.5	10.3	312	2 T35413	probable secreted
23	169	10.1	223	2 F85941	conserved lipoprote
24	167	10.0	371	2 A71359	conserved hypochet
25	158.5	9.5	436	2 AH1387	cell wall binding
26	158	9.5	375	2 F75467	probable lipoprote
27	155	9.3	437	2 AB1763	cell wall binding
28	153	9.2	228	2 T36379	probable peptidase
29	152	9.1	760	2 AB2225	hypothetical prote

30	146.5	8.8	439	2 H86658	N-acetyl[muramoyl]-L
31	146.5	8.8	482	2 AG1147	P60 extracellular
32	146.5	8.8	484	2 A41487	protein P60 precu
33	139.5	8.4	465	2 AG1506	P60 extracellular
34	138.5	8.3	715	2 S76492	lipoprotein nlpD -
35	136.5	8.2	285	2 C82736	hypothetical prote
36	135	8.1	420	2 D64158	hypothetical prote
37	134	8.0	428	2 A83005	conserved hypochet
38	132	7.9	475	2 F64151	hypothetical prote
39	131	7.8	376	2 S71558	probable cell wall
40	130.5	7.8	665	2 E75461	probable cell wall
41	130.5	7.8	894	2 C86756	prophage p12 prote
42	129.5	7.8	284	2 B84118	stage II sporulati
43	129	7.7	1711	2 T31337	1,4-Beta-glucanase
44	128.5	7.7	417	2 F70132	conserved hypochet
45	126.5	7.6	374	1 LYXIX	beta-lytic metallo

ALIGNMENTS

RESULT 1
S55063
conserved hypothetical protein PA3623 [imported] - Pseudomonas aeruginosa
C/Species: Pseudomonas aeruginosa
C/Date: 27-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 31-Dec-2000
C/Accession: S55063; E83193
R/Tanaka, K.; Takahashi, H.
Gene 150, 81-85, 1994
A/Title: Cloning, analysis and expression of an ipos homolog gene from Pseudomonas aer
A/Reference number: S55062; MUID:95047554; PMID:7959068
A/Accession: S55063
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-297 <TAN>
R/Seover, C.K.; Pham, X.Q.; Erwin, A.L.; Mitoguchi, S.D.; Warren, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lardig, K.; Lim,
; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A/Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathog
A/Reference number: A82950; MUID:20437337; PMID:10984043
A/Accession: E83193
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-297 <STO>
A/Cross-references: GB:AE004782; GB:AE004091; NID:99494772; PID:AG07011.1; GSPDB:GN001
A/Experimental source: strain PA01
C/Genetics:
A/Gene: PA3623
C/Superfamily: lipoprotein D
Query Match 17.8%; Score 298; DB 2; Length 297;
Best Local Similarity 28.0%; Pred. No. 2.7e-14;
Matches 91; Conservative 50; Mismatches 106; Indels 78; Gaps 12;

QY 20 LITGVITTCILACASKP-----TNTSSGSGSHRTSSGGLAIGQVITTSQGVPNR 72
Db LLLGAVVCSLLAACSSSPGVGVKVDNKGSAAPAARTPTTSG-----Q 68
QY 73 YQVKGQDTVSKIAQRYGLANREIGHINNLSYTIYTGWLTLMSGDLKVERSSISGVN 132
Db YIIRRGDTLXSLAFRQMDKALAAANGIAFPPTIQVQ----- 107
QY 133 TAHTPSFVAVQ--SSRPVQGHVAVQKPTPPV--VVKKPTPTPPVQO-----QPAVAPP 184
Db 108 -----AIOFGGRAGSTQ--PSVAKVTPVVAAPVATKPTPVPAVSTSVAPAPAP-A 156
QY 185 VTEAPATGSSGMQO-----FRYPVCAITPVVRRGTATVAVASTYTSNQMFTSSGRGDL 238
Db 157 STTPPSSGATPVVAGVAGVGMWAPASGT--LIGRF-----ASNGSLINKGIDIDAGQIGOP 209
QY 239 INASNAGTVIQADHNDGAS--IVIQHTNGFVSSYIIHKDAQVKTGDTVATGQRIASMKN 296

Db 210 VLAASGTTVVAAGSLRGGELVITIKHNTVSAVGNHRLVLEGQVKKVGSIAIMGS 269
 Qy 297 QPSGAALFEFRISRNQVVDPLTVL 331
 Db 270 TGTDRVKLHFEIRROGKRPVDPLOYL 294

RESULT 2

hypotheetical protein nlpd [imported] - Rubrivivax gelatinosus
 C/Species: Rubrivivax gelatinosus
 C/Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 21-Jul-2000
 C/Accession: T50918
 R/Nagashima, K.V.; Igarashi, N.; Harada, J.; Nagashima, S.; Matsunura, K.; Shimada, K.
 submitted to the EMBL Data Library, November 1999
 A/Description: Determination of Nucleotide Sequences of Rubrivivax gelatinosus Photosynth
 A/Reference number: 225270
 A/Accession: T50918
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 314 <NAC>
 A/Cross-references: EMBL:AB034704; PDB:BA94071.1
 A/Experimental source: strain IL14
 A/Genes: nlpd

Query Match 16.0%; Score 268; DB 2; Length 314;
 Best Local Similarity 25.6%; Pred. No. 4.2e-12;
 Matches 84; Conservative 43; Mismatches 161; Indels 40; Gaps 8;
 Qy 10 QNOKPRL---GLIFGVTTCILAGCASKPTYNSTSGSGSHRTSGGLAIGSVI-- 63
 Db 8 ENSRPGTRLRPAPSLILMAVAALVAGCAN-PSHRAPEVDRSGRPVASSPASAPETAK 66
 Qy 64 ----TDSQGVNRYQVQKQDVTASKIAQRYGLMREIGHINNLSSTIYTGQMLTMSGD 119
 Db 67 PPPGQENLGRPGYVAVPGDTLIRIGLENGQWRDIAFWNNLENPRILEVQV----- 120
 Qy 120 LKVERSSISGVN---TAHTPSFAVVOSSRPVQGHAPVOKPTPPVAVVKKPTPPV 175
 Db 121 ----RVPPGADPNGVVRPVAPARASRPASSGASASAPAPAAAT--PAAASTAS 172
 Qy 176 QCPAPVAPVTEAPFATGSGVWQGFYVPGATNPVVRRCZATVAGSTYTSNGMFGSD 235
 Db 173 ASPASGSSSSSVPAARDGDDVMTWTPAA-----GSVYAGFDEVRSGKLSIAGKA 223
 Qy 236 GDLINSMNGTVIOADHNMDSG--IVIOHTNGFVSSYTHIDAQVKTGDTVTRTGRIAS 293
 Db 224 GPPVVAADGVRVYVAGSGIRGNLTIIVGNATFSAVAHNOGTLIVKEDQAVRQOKIAE 283
 Qy 294 MNKQPSGALFEFRISRNQVVDPLTVL 331
 Db 284 MGASDADRQVLFELIRQKRPVDPKXL 311

RESULT 3

lipoprotein XPO855 [imported] - Xylella fastidiosa (strain 9a5c)
 A/Accession: A82754
 C/Species: Xylella fastidiosa
 C/Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
 C/Accession: A82754
 R/Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
 A/Title: The genome sequence of the plant pathogen Xylella fastidiosa.
 A/Note: For a complete list of authors see reference number A59328 below
 A/Accession: A82754
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-242 <SIM>
 A/Cross-references: GB:AE003925; GB:AE003849; NID:g9105750; PDB:AAF8365.1; GSPDB:GN001
 A/Experimental source: strain 9a5c

R/Stimpson, A.J.G.; Reimach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; Al
 brunes, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carver, H
 de-Neto, E.; Docena, C.; El-Dorri, H.; Faciniant, A.P.; Ferreira, A.J.S.
 submitted to GenBank, June 2000
 A/Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franco, S.C.; Franco, M.C.; Frohme
 J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kurame, E.E.; Laigre
 chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins E
 A/Authors: Martins, E.M.F.; Matsumura, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.
 F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmeri, D.A.
 Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasaki
 M.; Teuhako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
 A/Reference number: A59328
 A/Contents: annotation
 C/Genetics:
 A/Genes: XPO855

Query Match 15.1%; Score 253; DB 2; Length 242;
 Best Local Similarity 28.5%; Pred. No. 3.6e-11;
 Matches 87; Conservative 32; Mismatches 108; Indels 78; Gaps 9;
 Qy 20 LIFGVTTCILAGCASKPTYNSTSGSGSHRTSGGLAIGSVITTSQGVNR-YQVKOG 78
 Db 10 LLSTALALALACSTATTVPNNNTSKRTA-----KPNQTVVKKOG 52
 Qy 79 DTASKIAQRYGLMREIGHINNLSSTIYTGQMLTMSGD LKVERSSISGVNHTPS 138
 Db 53 DTLVAISRRTGVAPQDLAMNRULPASKTIPPGVLRLEPDD----- 93
 Qy 139 PVAOSSRPVQGHAPVOKPTPPVAVVKKPTPPVQVQCPAPVAPVTEAPATSSGVW 198
 Db 94 ----ATASPTPQPTSPPTSPENPPTTITIAIPANSGN 128
 Qy 199 QFRYPVATNPVVRREGTATVAGSTYTSNGMFGSDGDLINSMNGTVIOADHNMDSG 258
 Db 129 WFPETGA---VVSNF---VAGOT-TKQGVSTNGNNGGTITIAANGTVYVGSALIGY 180
 Qy 259 --IVIOHTNGFVSSYTHIDAQVKTGDTVTRTGRIASMNKQPSGALFEFRISRNQVVD 316
 Db 181 ELIIKINEGIMISAYGNHNRRLVNEGQIVKANQPIEM-----GMLTY-FEIRYNGKPYD 234
 Qy 317 PLTVL 331
 Db 235 PLTVL 239

RESULT 4

hypotheetical protein b2865 - Escherichia coli (strain K-12)
 C/Species: Escherichia coli
 C/Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
 C/Accession: A65070
 R/Battner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Col
 science 277, 1453-1462, 1997
 A/Title: The complete genome sequence of Escherichia coli K-12.
 A/Reference number: A64720; PMID:97426517; PMID:9278503
 A/Accession: A65070
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown
 A/Molecule type: DNA
 A/Residues: 1-259 <BIAT>
 A/Cross-references: GB:AE000369; GB:U00096; NID:g2367168; PDB:AACT5903.1; PID:g1789228;
 A/Experimental source: strain K-12, substrain MG1655
 C/Superfamily: lipoprotein D

Query Match 14.3%; Score 238; DB 2; Length 259;
 Best Local Similarity 24.9%; Pred. No. 4.8e-10;
 Matches 78; Conservative 46; Mismatches 103; Indels 86; Gaps 8;
 Qy 16 KRLGLIFVITTCILAGCASKPTYNSTSGSGSHRTSGGLAIGSVITTSQGVNRYQV 75
 Db 17 KSLGIIVMLSVGLLHAGC-----SGSKSDPTGYSG-----SYTVT 52

QY 76 KQDVTAKIQRVGLNREIGHINNLSSTYITYGQWLTMSGDLK---VRESSISGV 131
 Db 53 KRGTLYKISRTTSTSVLELRLNGLISPPYITVEGQKRLK---GGAKSSSIRKSTAKSTT 110
 QY 132 NTAA-TTSPVAVQSSRPVQOHPAVOKTTPPVVVKKTPPTPPVQOQAPAPAPPTTEAPF 190
 Db 111 KTAATVTPSSAVPKSSWPPVQRCWLMPTTGKVM-----PY 146
 QY 191 ATGSSGVQAFRYPVGATNPVRRFGTATVAGSTVTSNGMFPSSGRDGLINASNAGTVIA 250
 Db 147 STADGG-----NKGIDISAPRGFTIYAAGKAVYV 177
 QY 251 DHNMDSAS--IVIOHTNGFVSSYIHIKDAQVKTGDTVTRGRIASMKQPSGALLFEFRI 308
 Db 178 GNGRGRGNLIMIKHSEDEYITAYAHNDTMLVNNQGSVAGQKIATMGSTDAASVRLHFOI 237
 QY 309 SRNGVYVDPPLTVL 321
 Db 238 RYRATATDPLRYL 250

RESULT 5

B91096
 Probable lipoprotein [imported] - Escherichia coli (strain O157:H7, substrain RIMD 05093
 C/Species: Escherichia coli
 C/Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 14-Dec-2001
 C/Accession: B91096
 R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;
 Gotoh, N.; Yasunaga, T.; Kihara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
 DNA Res. 8, 11-22, 2001
 A/Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene
 A/Reference number: A99629; MUID:21156231; PMID:11258796
 A/Accession: B91096
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-259 <HAY>
 A/Cross-references: GB:BA000007, PIDN:BA037161.1, PID:G13363210; GSPDB:GN00154
 A/Experimental source: strain O157:H7, substrain RIMD 0509352
 C/Genetics:
 A/Gene: EC03738
 C/Superfamily: lipoprotein D

Query Match 14.3%; Score 238; DB 2; Length 259;
 Best Local Similarity 24.9%; Pred. No. 4.8e-10;
 Matches 78; Conservative 46; Mismatches 103; Indels 86; Gaps 8;

QY 16 KRLLIFGVITTCILAGCASKPTNSTSGSGSHRTSGSGGLAISQVITDSQVFNRYQV 75
 Db 17 KSLGIWMLSVGLLAGC-----SGSKSPDTGYSG-----SVYTV 52
 QY 76 KQDVTAKIQRVGLNREIGHINNLSSTYITYGQWLTMSGDLK---VRESSISGV 131
 Db 53 KRGTLYKISRTTSTSVLELRLNGLISPPYITVEGQKRLK---GGAKSSSIRKSTAKSTT 110
 QY 132 NTAA-TTSPVAVQSSRPVQOHPAVOKTTPPVVVKKTPPTPPVQOQAPAPAPPTTEAPF 190
 Db 111 KTAATVTPSSAVPKSSWPPVQRCWLMPTTGKVM-----PY 146
 QY 191 ATGSSGVQAFRYPVGATNPVRRFGTATVAGSTVTSNGMFPSSGRDGLINASNAGTVIA 250
 Db 147 STADGG-----NKGIDISAPRGFTIYAAGKAVYV 177
 QY 251 DHNMDSAS--IVIOHTNGFVSSYIHIKDAQVKTGDTVTRGRIASMKQPSGALLFEFRI 308
 Db 178 GNGRGRGNLIMIKHSEDEYITAYAHNDTMLVNNQGSVAGQKIATMGSTDAASVRLHFOI 237
 QY 309 SRNGVYVDPPLTVL 321
 Db 238 RYRATATDPLRYL 250

RESULT 6

B83031

conserved hypothetical protein PA4924 [imported] - Pseudomonas aeruginosa (strain PA01)
 C/Species: Pseudomonas aeruginosa
 C/Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
 C/Accession: B83031
 R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Br
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lapidig, K.; Lim,
 .; Lory, S.; Olson, M.V.
 Nature 406, 959-964, 2000
 A/Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
 A/Reference number: A82950; MUID:20437337; PMID:10958403
 A/Accession: B83031
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-231 <STO>
 A/Cross-references: GB:AE004905; GB:AE004091, NID:9951195; PIDN:AA080309.1; GSPDB:GN001
 A/Experimental source: strain PA01
 C/Genetics:
 A/Gene: PA4924

Query Match 14.0%; Score 234; DB 2; Length 231;
 Best Local Similarity 26.7%; Pred. No. 8e-10;
 Matches 68; Conservative 37; Mismatches 90; Indels 60; Gaps 6;

QY 69 VPNEYQVQDVTAKIQRVGLNREIGHINNLSSTYITYGQWLTMSGDLKVERSSIS 128
 Db 32 VRGEYRVKRGDTLYSIATRHGMNYKDLARANGIRPVAVKQGVREPDGRKSTYVASSR 90
 QY 129 SGVATATPSPVAVQSSRPVQOHPAVOKTTPPVVVKKTPPTPPVQOQAPAPAPPTTEA 188
 Db 91 SSSNTR-----ARKPPPPPSVT----- 108
 QY 189 PFATSSGVQAFRYPVGATNPVRRFGTATVAGSTVTSNGMFPSSGRDGLINASNAGTVIA 248
 Db 109 -----LKGWQPM--KEPVIRF-----SSDKNKIRIGTIGQPVQASLAGRV 153
 QY 249 QADHNDGAS--IVIOHTNGFVSSYIHIKDAQVKTGDTVTRGRIASMKQPSGALLFEF 306
 Db 154 FAVNMGYGNLVIIOHTGTSYTSYANSLRLVYEGWVGQKIAEAGSSDADRVLQYF 213
 QY 307 RISNGVYVDPPLTVL 321
 Db 214 EIRONGRPLDPLSL 228

RESULT 7

AD0871

Probable lipoprotein [imported] - Salmonella enterica subsp. enterica serovar Typhi (str
 C/Species: Salmonella enterica subsp. enterica serovar Typhi
 A/Note: this species has also been called Salmonella typhi
 C/Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 27-Nov-2001
 C/Accession: AD0871
 R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
 th, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
 .; Moul, S.; O'Gaora, P.
 Nature 413, 848-852, 2001
 A/Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
 A/Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
 A/Reference number: AB0502; PMID:11677608
 A/Accession: AD0871
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-250 <PAR>
 A/Cross-references: GB:AL513382; PIDN:CAP02868.1; PID:G16504121; GSPDB:GN00176
 C/Genetics:
 A/Gene: STY3194
 C/Superfamily: lipoprotein D

Query Match 13.9%; Score 232.5; DB 2; Length 250;
 Best Local Similarity 24.4%; Pred. No. 1.1e-09;
 Matches 76; Conservative 49; Mismatches 99; Indels 87; Gaps 9;

QY 18 LGLIFGVITTCILAGCASKPTNSTSGSGSHRTSGSGGLAISQVITDSQGVFNRYQVQ 77
 Db 17 KSLGIWMLSVGLLAGC-----SGSKSPDTGYSG-----SVYTV 52

Db 11 LGIAVLCTGILLAGC-----SSNSGSGTY--SGS-----VYTVKR 44
Qy 78 GDYVSKIAQRYGLNMRREIGHINNINSSTIYTGOMLTMSGDLK-----VERISISGVT 133
Db 45 GDTLYRISRATGTGSKELARLNGISPPYTIIEVGRIKV-RGSAASSSTKTSKTKTKT 103
Qy 134 AHT-PSPVAVQSSRPVQOHPAVQKPTPPVAVVKKPTPTPPVQGPAPVAPVTEAPAT 192
Db 104 AAAPVSSSSVPKSSVPVQRCWVPWANGVIL-----PYST 139
Qy 193 GSSGVMOFRYPVATNPVAFGTAVAGSTVTSNGMWSGSGRQDGLINASNACTVIOADH 252
Db 140 AEGG-----NKGDIDAGAGTPTVYAGAGKVVAVGN 170
Qy 253 NMDGAS--IYIOTHTNGVSSYIHKDAQVKTGPTVRCGRIASMKQPSGALFEPRISR 310
Db 171 QLGKYGULIMIKNKEDITTAHANDTMLVANNQSVKAGOKIATAGSTDAASVRLHFQIRY 230
Qy 311 NGYVVDPLTVL 321
Db 231 RATAIDPLRYL 241

RESULT 8

lipoprotein D homolog - Haemophilus influenzae (strain Rd KW20)
C/Species: Haemophilus influenzae
C/Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 29-Sep-1999
C/Accession: F64087; T09410
R/Fischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, J.; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, J.M.; Weidman, J.; D.M.; Brandon, R.C.; Fine, L.D.; Fitchman, J.L.; Fuhmann, J.L.; Geoghegan, N.S.M. Science 269, 496-512, 1995
A/Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, A./Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A/Reference number: A64000; MID:95350630; PMID:7542800
A/Accession: F64087
A/Molecule type: DNA
A/Status: nucleic acid sequence not shown; translation not shown
A/Residues: 1-405 <TRIG>
A/Cross-references: GB:U32753; GB:LA2023; NID:91573701; PIDN:AAC2363.1; PID:91573707; T
A/Genetics:
A/Gene: H10706
C/Superfamily: lipoprotein D

Query Match 13.1%; Score 219.5; DB 2; Length 405;
Best Local Similarity 27.8%; Pred. No. 1.8e-08;
Matches 76; Conservative 46; Mismatches 114; Indels 37; Gaps 10;

Qy 71 NRYVQKQGTGSKIAQRYGLNMRREIGHINNINSSTIYTGOMLTMSGDLKVRERSISSG 130
Db 145 NRYVQKQGTGSKIAQRYGLNMRREIGHINNINSSTIYTGOMLTMSGDLKVRERSISSG 130
Qy 131 VNTAHTPSFVAVQSSRPVQOHPAVQKPTP-----PVVVVKKPTPTPPV 175
Db 203 -----VKQAVTASATATPK--PAVTV--TPGANGTQISDGTIIIPISKAGTSFVVA 254
Qy 176 QGPAPVAPVTEAPATGSSGVMOFRYPVATNPVAFGTAVAGSTVTSNGMWSGSGRQDGLINASNACTVIOADH 252
Db 255 TSSTVTSVANNANSTPTNSNVA--PI-ASN-VVWOMPTSGNIIOGFSSTDSGNGKID 309
Qy 231 FSGRQDGLINASNACTVIOADHNDGAS--IYIOTHTNGVSSYIHKDAQVKTGPTVRCGRIASMKQPSGALFEPRISR 310
Db 310 ISGSGCAVVKAAARIVAGNALRGYGNLIIKKNDDFLAVYAHNDKIIVADQGVKAG 369
Qy 289 QRIASMKQPSGALFEPRISRNGYVVDPLTVL 321
Db 370 QDIAMGSSGNTNTVGLHFEIRYKSGVDPVRYL 402

RESULT 9
E75291
probable cell wall glycyL-glycine endopeptidase - Deinococcus radiodurans (strain R1)

C/Species: Deinococcus radiodurans
C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C/Accession: E75291
R/White, O.; Eilen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; F.M.; Shen, M.; Vamathevan, J.U.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma Science 286, 1571-1577, 1999
A/Title: Genome sequence of the radioreistant bacterium Deinococcus radiodurans R1.
A/Reference number: A75250; MID:20036896; PMID:10567266
A/Accession: E75291
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-454 <WHI>
A/Cross-references: GB:AE002061; GB:AE000513; NID:96460095; PIDN:AF11838.1; PID:9646010
A/Experimental source: strain R1
A/Genetics:
A/Gene: DR2291
A/Map position: 1

Query Match 13.1%; Score 219.5; DB 2; Length 454;
Best Local Similarity 26.1%; Pred. No. 2.1e-08;
Matches 73; Conservative 50; Mismatches 116; Indels 41; Gaps 9;

Qy 60 SOVITDSGVNRYVQKQGTGSKIAQRYGLNMRREIGHINNINSSTIYTGOMLTMSGDL 119
Db 187 AQLAAKKRPTTHRVIGDTFYVAKRIGINPLADEVPRILGQTLNAGVLSVAPP 246
Qy 120 LKVERISISGVTNHTPPSPVAVQSSRP-PVQ--QHPAVQKPTPPVAVVKKPTPTPPV 175
Db 247 LR-----FAVPA--PAKPAVRLVAPAPASPTPAPAPAPAPAPAP 289
Qy 176 QGPAPVAPVTEAPAT-----GSSGVMOFRYPVATNPVAFGTAVAGSTVY 225
Db 290 ARKPAQAPAPPAVQATPEHHARVAVVQTSLSLWQMPLEGYRITSDGWRVLDGEREK 349
Qy 226 SNGMWSGSGRQDGLINASNACTVIOADHND-----GASIVOTHTNGVSSYIHKDAQV 280
Db 350 HQGIDVAPPGTPTVIAARSSRVIOA-HIDETYGWMTVVIQHPDQMQRVHLSRISVE 407
Qy 281 TGDIVRTGRIASM--KNQPSGALFEPRISRNGYVVDPL 318
Db 408 AGQLVROGERVAVGSGTGRVTPHLL-HFGLYRMWDPHNP 446

RESULT 10

lipoprotein D [imported] - Yersinia pestis (strain CO92)

C/Species: Yersinia pestis
C/Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 17-May-2002
C/Accession: AF0407
R/Parhelli, U.; Wren, B.W.; Thomson, N.R.; Tibball, R.W.; Holden, M.T.G.; Prentice, M.B.; deno-Tarraga, A.M.; Chillingworth, T.; Crofton, A.; Davies, R.M.; Dougan, G.; F. Nature 413, 523-527, 2001
A/Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A/Reference number: AB0001; MID:21470413; PMID:11586360
A/Accession: AF0407
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1333 <KOR>
A/Cross-references: GB:AL590842; PIDN:CAC92586.1; PID:915981283; GSPDB:GN00175
A/Genetics:
A/Gene: nlpD
C/Superfamily: lipoprotein D

Query Match 13.1%; Score 218.5; DB 2; Length 333;
Best Local Similarity 25.3%; Pred. No. 1.7e-08;
Matches 91; Conservative 52; Mismatches 130; Indels 87; Gaps 14;

Qy 14 PIKRIQLIFG-VITTCIILAGCASKRPTVSTSGSGHRTSGGGLAIG--SOVITDSGV 68
Db 6 PMIRLRVAACTVAVLMLVGTND--NSTSAP-----ISSVGDRSGTMLSKANIDSOG 58

Or 220 VT 221

A;Experimental source: strain 16M

C/Genetics:

C/Accession: D91078
 R/Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;
 Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
 DNA Res. 8, 11-22, 2001
 A/Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genc
 A/Reference number: A99629; MUID:21156231; PMID:11258796
 A/Accession: D91078
 A/Status: Preliminary
 A/Molecule type: DNA
 A/Residues: 1-379 <HAV>
 A/Cross-references: GB:BA00007; PIDN:BA37019.1; PID:g13363067; GSPDB:GN00154
 A/Experimental source: strain O157:H7, substrain R1MD 0509952
 C/Genetics:
 A/Gene: EC63596
 C/Superfamily: Lipoprotein D

Query Match 12.1%; Score 202.5; DB 2; Length 379;
 Best Local Similarity 26.1%; Pred. No. 2.8e-07;
 Matches 72; Conservative 42; Mismatches 113; Indels 49; Gaps 9;

Qy 73 YVQKQDVTAKIARVGLNREIGHINNLSYITYGQMLTWSGDLKVERISISGVN 132
 Db 123 YTVKKGDTLFYIANTGNDPFDLQKRNITQAPYALNNGQ--TLQVG-----N 167
 Qy 133 TAHTPSPVAVQSSRPVQOHPAVQKP--TPVVVVVKKPT-----P 170
 Db 168 ASGPTPIGNAITQADAEGGVVKKPAQNSTVAVASQPTIYSSSGEGSANKMLPNKKP 227
 Qy 171 TTPVVOQAPAPVAPVTEAPFATG---SSGVMOFRYPVGAIPVVRREGTATVAGSTVTSN 227
 Db 228 TATVTLPVVPVPTASTEPVTSSTSTSTPSTWRP--TEGKVIETFG--ASEGG---NK 280
 Qy 228 GMMFSGRBDGLINASNAGTVIQADHNDGAS--IVIOHTNGFVSSYHIKDAQVKTGDTV 285
 Db 281 GIDTAGSKGAIIATADGRVYAGNALRGYNLIIIKINDDYLSAVYANDTMLVREQQEV 340
 Qy 286 RTGORIASMKQPSGALFFERISRNNGVYDPLTVL 321
 Db 341 KAGOKIATMGSTGTSTRHFEIRYKGSVNPRLYL 376

RESULT 17
 E85923
 Lipoprotein [imported] - *Escherichia coli* (strain O157:H7, substrain EDL933)
 C/Species: *Escherichia coli*
 C/Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 17-May-2002
 C/Accession: E85923
 R/Perma, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
 Miller, L.; Grobeck, B.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
 Nature 409, 529-533, 2001
 A/Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.
 A/Reference number: A85460; MUID:21074935; PMID:11206551
 A/Accession: E85923
 A/Status: Preliminary
 A/Molecule type: DNA
 A/Residues: 1-379 <STO>
 A/Cross-references: GB:AE005174; NID:g12517196; PIDN:AA657849.1; GSPDB:GN00145; UMGF:Z4C
 A/Experimental source: strain O157:H7, substrain EDL933
 C/Genetics:
 A/Gene: nlpD
 C/Superfamily: Lipoprotein D

Query Match 12.1%; Score 202.5; DB 2; Length 379;
 Best Local Similarity 26.1%; Pred. No. 2.8e-07;
 Matches 72; Conservative 42; Mismatches 113; Indels 49; Gaps 9;

Qy 73 YVQKQDVTAKIARVGLNREIGHINNLSYITYGQMLTWSGDLKVERISISGVN 132
 Db 123 YTVKKGDTLFYIANTGNDPFDLQKRNITQAPYALNNGQ--TLQVG-----N 167
 Qy 133 TAHTPSPVAVQSSRPVQOHPAVQKP--TPVVVVVKKPT-----P 170
 Db 168 ASGPTPIGNAITQADAEGGVVKKPAQNSTVAVASQPTIYSSSGEGSANKMLPNKKP 227

Qy 171 TTPVVOQAPAPVAPVTEAPFATG---SSGVMOFRYPVGAIPVVRREGTATVAGSTVTSN 227
 Db 228 TATVTLPVVPVPTASTEPVTSSTSTSTPSTWRP--TEGKVIETFG--ASEGG---NK 280
 Qy 228 GMMFSGRBDGLINASNAGTVIQADHNDGAS--IVIOHTNGFVSSYHIKDAQVKTGDTV 285
 Db 281 GIDTAGSKGAIIATADGRVYAGNALRGYNLIIIKINDDYLSAVYANDTMLVREQQEV 340
 Qy 286 RTGORIASMKQPSGALFFERISRNNGVYDPLTVL 321
 Db 341 KAGOKIATMGSTGTSTRHFEIRYKGSVNPRLYL 376

RESULT 18
 H81079
 Lipoprotein NlpD, probable NMB1483 [imported] - *Neisseria meningitidis* (strain MC58 serog
 C/Species: *Neisseria meningitidis*
 C/Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
 C/Accession: H81079
 R/Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A
 Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;
 Ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizzza, M.
 Science 287, 1809-1815, 2000
 A/Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ver
 A/Title: Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58.
 A/Reference number: A81000; MUID:20175755; PMID:10710307
 A/Accession: H81079
 A/Status: Preliminary
 A/Molecule type: DNA
 A/Residues: 1-415 <TEPT>
 A/Cross-references: GB:AE002497; GB:AE002098; NID:g7226712; PIDN:AA41839.1; PID:g722672;
 A/Experimental source: serogroup B, strain MC58
 C/Genetics:
 A/Gene: NMB1483

Query Match 11.9%; Score 198; DB 2; Length 415;
 Best Local Similarity 27.2%; Pred. No. 6.8e-07;
 Matches 72; Conservative 38; Mismatches 95; Indels 60; Gaps 12;

Qy 73 YVQKQDVTAKIARVGLNREIGHINNLSYITYGQMLTWSGDLKVERISISGVN 130
 Db 185 HTIVRGDTVYNISIKRY-----HISQD-----FRWNG---MTDNLSIGQI 223
 Qy 131 --VNTATPSP--VAVQSSRPVQOHPAVQKPTPPVVKPTPPVVOQAPAPAP---- 183
 Db 224 VKVPAQYAAPKTAAVESRPAPV--AAVQTPV-----TPAAPVQSAPOPAAPAAEN 274
 Qy 184 -----PVTEAPFA-----TGSSGVMOFRYPVGAIPVVRREGTATVAGSTVTSNGMFSGR 234
 Db 275 KAVPAPAPQSPAPSPGSTRVGVGVQRP--TGKVVADGCG-----NNKGVDIAGN 324
 Qy 235 DGDGLINASNAGTVIQADHNDGAS--IVIOHTNGFVSSYHIKDAQVKTGDTVFGRIA 292
 Db 325 AGQPVLAADGKVVYAGSGLRGVGNLVIIOHNSFLTAAGHNOGLVGEQGVVRGQOVA 384
 Qy 293 SMKQPSGALFFERISRNNGVYDPL 317
 Db 385 LMGNTASRTQLHFEVRQNGKVPNP 409

RESULT 19
 G87496
 Peptidase, M23/M37 family [imported] - *Caulobacter crescentus*
 C/Species: *Caulobacter crescentus*
 C/Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
 C/Accession: G87496
 R/Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.H
 B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolonja
 n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
 Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
 A/Title: Complete Genome Sequence of *Caulobacter crescentus*.
 A/Reference number: A87249; MUID:21173698; PMID:11259647

A:Accession: G87496
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-609 <SPO>
A:Cross-references: GB:AE005673; NID:gl3423465; FID:AAK2371.1; GSPDB:GN00148
C:Genetics
A:Gene: CCI1996

Query Match	11.7%	Score 196;	DB 2;	Length 609;
Best Local Similarity	23.6%	Pred. No. 1.5e-06;		
Matches	66;	Conservative	54;	Mismatches 110;
			Indels	50;
			Gaps	12

```

QY      73  YQVKGQDPIVSKIAQGYGLNWEIGHINNLSYTYTQOMTLWGGDLKVERSISSQVN 132
Db      347  YSVQGGDTLGEIARKEFNVSVKALAEENLRLATASLKKQKIALPDGG---FRDK--GPIR 400

QY      133  TAHTSSPAVQSSRPVQOHPAVQKPTPPVVVVKKPTPTPPVQOAPAVP---PYEA 188
Db      401  TTTTRP-----ATRPANTYARVDSASAAASPSSPVYTPSGAFAFRSAPFAVAQPIINP 455

QY      189  PPATG-----SSGVMOFRYPGATNPVVRRGATVAGSTVTSQMMFSS 232
Db      456  P-SSORTIITETAAPTAEAILIASKGKFAWPL--RGDIISGG---VAGTQGRNDGLNIR 509

QY      233  GRDGLLIANSNGYITQADHNND--GASIVIOHTMGFVSSTYHIMDAQVKTGDTYRTGQR 290
Db      510  AFQGFPIVLSADGEIAYAKNOVPTFGNVLVLAHDAQOWTAYAHLSSTNVKRGQVKGQSEQ 569

QY      291  IASM-----KQPSGALPEFR---ISRGYVNDPLIVL 321
Db      570  LGIVGATGAVNEPQ--LHEFWRYAPTVKDKAKPFAVLVL 607

```

RESULT 20
AF2785
lipoprotein [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C.Species: Agrobacterium tumefaciens
C.Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
C.Accession: AF2785
R.Wood, D.W. Stethal, J.C. Kaut, R. J. Monks, D. J. Chen, L. J. Wood, G.E. J. Chen, Y. J. Woo, L. J. Stethal, G. J. Gallet, W. J. Grant, C. J. Guenther, D. J. Kuyavlin, T. J. Levy, R. J. Li, M. J. McClellan, P. J. Romero, P. J. Zhang, S.
Science 294, 2317-2323, 2001
A.Authors: Yoo, H., Teo, Y., Biddle, P., Jung, M., Krespan, W., Perry, M., Gordon-Kamm, A., E.W. Steier
A.Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A.Reference number: AB2577, PMID:11743193
A.Accession: AF2785
A.Status: Preliminary
A.Molecule type: DNA
A.Residues: 1-537 <NOR>
A.Cross-References: GB:AE008688; PIND:AA142700.1; PID:g17740136; GSPDB:GN00186
A.Experimental source: strain C58 (Dupont)
C.Genetics:
A.Gene: Atu1700
A.Map position: circular chromosome

[illegible]

QY 18 VAPPVTEAFATGSSGWAQFRTYV-GAATNPVRRFGATATAGSVTNSGNGMFSGRGDILI 239

Db 398 IEKKADMASTAPESGTIGKRWPVGCA--VINNFGD-NEYGS--RNDGGINIVPEKPTPI 451

QY 240 NAAATGTVIQAQHNMD--GASIVIOHTGTFPVSSTYTHIKDAQVKTGDTVFRIGCIATSMKQ 297

Db 452 KAALENGVVVLYAAGNGKQLQANTVLVYVHDDGKVTYVGAANALNDVQGRQVQRGQTIAT--SG 509

QY 298 PSGAA---LFEFRISNNGVYDP 317

Db 510 MTGSARKPQVHFEVRRDAPVPV 532

```

RESULT 21
H97564
43K antigen (AF157831) [imported] - Agrobacterium tumefaciens (strain C58, Cereon)
C.Species: Agrobacterium tumefaciens
C.Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jan-2002
C.Accession: H97564
R.Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Girollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allington, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A.Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens
A.Reference number: A97359; PMID:11743194
A.Accession: H97564
A.Status: preliminary
A.Molecule type: DNA
A.Restrases: 1-562 (KHF-
A.Cross-references: GB:AE007869; PIDN:AAK87473.1; PID:gl5156795; GSPDB:GN00169
C.Genetics:
A.Gene: AGR_C_3124
A.Map position: circular chromosome

```

Query Match	11.3%	Score 188.5	DB 2	Length 562
Best local similarity	25.1%	Pred. No. 4.8e-06		
Matches	81	Conservative	64	Mismatches 131; Indels 47; Gaps 15

QY	5	IAINQONOKPIKRLGLIGVITTCITLACACASKPTVNSTSGSGSHNTSGSGGIA-IGSQVI	63
DB	272	IALNNGNQF-----TFLRADGNVAVALPQAAARDKUSSEAGKTPPGGKPL	319
QY	64	TDGCGVPRNRXVYKQSGDTSKTAQRGLMWREIHHNNINSSSTIYTGWLTLMSGDLKVR	123
DB	320	PPSGG----YKVPDSDLAKIARAGVSAALKANGI-SNRSINVGQTILM-----	366
QY	124	ERSISGVNTHATSEFVAVQSSRPVQOCHNAVQKPT---PVVVYKKEPTPRPVYVQOPAP	180
DB	367	PGASIDAIKITSVPKKA--AAKVE--TASIKREPKAPAAATTAAPATPATASVSD	422
QY	181	VAPPTAEAFATGSSGQWGFYFV-GAIPNVYARPGTATVAGSYTISNQMWSGSGDGI	239
DB	423	IEKADMASTIAEISGIGIKRHPVGA---VINNGD-NVBS--RNGININSVEGPTI	476
QY	240	NASNAGTYIQAHDHND--GASIVIGHTGFGVSYTHIDAQYKTDGYRTQTGRISMKNQ	297
DB	477	KLAENGVIYIAGNGIKQIAGNTVLVHHDDCKVTYIGNAANLDVYRQKQYRGGTIAF--SG	534
QY	298	PSGAA---LFEFRISRNGVYDP	317
DB	535	MTGSARPOVHFEVRKDALTPVNF	557

RESULT 22
T35413
probable secreted peptidase - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 05-Nov-1999 #sequence_rev:05-Nov-1999 #text_change 21-Jan-2000
C:Accession: T35413
R:Oliver, K., Harris, D., Bentley, S.D., Parkhill, J., Barrett, B.G., Rajandream, M.A.
submitted to the EMBL Data Library, March 1999
A:Reference number: Z21577
A:Accession: T35413
A:status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA
A:Residues: 1-312
A:Cross-references: EMBL:AL049485; PIDN: CAB39706.1; GSPDB: GN00070; SCOEDB: SC6A5.22
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOEDB: SC6A5.22
C:Superfamily: Lipoprotein D

Query Match 10.3%; Score 171.5; DB 2; Length 312;
Best Local Similarity 23.3%; Pred. No. 3.8e-05;
Matches 79; Conservative 43; Mismatches 110; Indels 99; Gaps 15;

```
Qy 46 GSHR-----TSGGGLAIGSVITDS-----QGV-----70
Dy 5 GKRRTRAMRLTRITAVGTGGAALALPLMGAAANAPASVSQAVQSPASAKXAA 64
Qy 71 -----NRQYKQDPTVSKIAQRYGLMREIGHINNNSSITTYTGQWLTMGDLKRE 124
Dy 65 EKNSDSRTYTKSGDYLSKIADBEDVD-----GGMKLYADN---R 102
Qy 125 RSISGVNTAHTPSPVAV--QSSRPVQHPAVOKPPTPVVVKKPTTPPVQOPAPVA 182
Dy 103 EAVSDPSLHPGLKLSIDGAAKAPASASQKPA-----QKSAKPAKTAAPKA 155
Qy 183 PPTVAPATGSSGVMQFRYPVGATNPVRRFGTATVAGSTV-----TSNGWFSG--- 233
Dy 156 QKSTDAKXASSDTGSQ---STGTTSGV-----TSPVAGTGVGPRPHQSSGMSGHTG 207
Qy 234 -----RQDGLINASNACTVIOAD--HNMDSIVIOHTNGFVSSYIHKDAQVKTGDTYRT 287
Dy 208 TDFVVPPTGSLKAVAGAGVVSAGWGAGVQVIOQLDGHVAYAHLSLSVSAGOSVTA 267
Qy 288 GORI--ASMKNQPSGAL--FEFRISRN--GVVVDPLTVLK 322
Dy 268 GQVGLSGATGATGVTGPHLHFEIRITTPDYGSDIDLAFUR 306
```

RESULT 23

probable lipoprotein 24203 [imported] - *Escherichia coli* (strain O157:H7, substrain EDLS F85941)
C:Species: *Escherichia coli*
C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Dec-2001
C:Accession: F85941
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001.
A:Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: F85941
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-223 <STO>
A:Cross-references: GB:AE005174; NID:912517383; PIDN: AAG57994.1; GSPDB: GN00145; UMG: Z42
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: 24203
C:Superfamily: Lipoprotein D

Query Match 10.1%; Score 169; DB 2; Length 223;
Best Local Similarity 24.0%; Pred. No. 3.8e-05;
Matches 63; Conservative 35; Mismatches 79; Indels 86; Gaps 8;

```
Qy 29 ILACASAFPTNSTRSGSHRTSSGGLAISQVITDSQVPMNVYQKQDPTVSKIAQRY 88
Dy 9 LLAGC-----SGKSDPTGYSG-----SVYTVRRGDTLIRISRTT 44
Qy 89 GLMREIGHINNNSSYTYTGQWLTMGDLK-----VRESISGVNTAH--TSPVAVQ 143
Dy 45 GTSVKEILARLNGISPPYITIEVGQKLK--GSAKSSSRKSTAKSTTKTATVTSNAVK 102
Qy 144 SSRPPVQHPAVOKPPTPVVVKKPTTPPVQOPAPVAPVTAEPATGSSGVMQFRYP 203
Dy 103 SSMPVPGRCWLWPTTGKVM-----PYSTADGG----- 131
```

```
Qy 204 VGATNPVRRFGTATVAGSTVTSNGWFSGRDGLINASNACTVIOADHNMDS--IVI 261
Dy 132 -----NKGIDISAPRGPIPTAAGAKVYVGNQLRGYNLIMI 169
```

```
Qy 262 QHTNGFVSSYIHKDAQVKTGDT 284
Dy 170 KHSEDYITAYAHNDTMLVNNQGS 192
```

RESULT 24

A:1359
conserved hypothetical protein TP0155 - *Syphilis spirochete*
C:Species: *Treponema pallidum* subsp. *pallidum* (*Syphilis spirochete*)
C>Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 05-Nov-1999
C:Accession: A71359
R:Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwin
rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Ueterbach, T.; McDo
ney, L.; Weidman, J.; Smith, H.O.; Venter, J.C.
Science 281, 375-388, 1998
A:Title: Complete genome sequence of *Treponema pallidum*, the *Syphilis spirochete*.
A:Reference number: A71359; MUID:9832770; PMID:9665876
A:Accession: A71359
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-371 <COL>
A:Cross-references: GB:AE001200; GB:AE000520; NID:93322419; PIDN: AAC65145.1; PID:9332242;
A:Experimental source: strain Nichols
C:Genetics:
A:Gene: TP0155

Query Match 10.0%; Score 167; DB 2; Length 371;
Best Local Similarity 24.9%; Pred. No. 0.0001;
Matches 72; Conservative 47; Mismatches 122; Indels 48; Gaps 11;

```
Qy 73 YQVQKQDPTVSKIAQRYGLMREIGHINNNSSYTYTGQWLTMGDLK-----LKYRERSISS 129
Dy 87 YEVEGQDVGRIAQRYDIOGDAIISLKLKSTALQVGLKIPSVDGILYTVNGDPTFS 146
Qy 130 GVNTHPSPVAVVQSSPPVQHPAVOKPPTPVV--VVKR-----TTPP----- 172
Dy 147 SIAAH-----QISLELVILNTPSSSKESPVRILTVPFYNASARESCVPPFPSSAKQ 201
Qy 173 -----PVQ--QPAVAPVTEAPATGSSGVMQ-----FRYPGATNPVRRFG-- 215
Dy 202 WRENTSDAVQIOPA-----RVLFPEAHLSARALQELINDLFRAPLRSTYVSSRYGMR 257
Qy 216 TATVAGSTVTSNGWFSGRDGLINASNACTVIOADH--MDGASIVIOHTNGFVSSYIHI 274
Dy 258 SDPFTGARSFHNGLDMVSRRTGYVSALGIVATVGSANVYGNVLIYGHAGYOTLXGHL 317
Qy 275 KDAVKTGDTVTRGQRIASW--KNQPSGAL--FEFRISRNQVYVVDPLTVLK 322
Dy 318 QTVLVSAQTVTSATKIGLIGKTRSGTGPMLHFTIYNAGSAINPTSLR 366
```

RESULT 25

AH1387
cell wall binding proteins homolog lmo2504 [imported] - *Listeria monocytogenes* (strain EC
C:Species: *Listeria monocytogenes*
C>Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
C:Accession: AH1387
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker,
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Duseurget, O.; Entlian, K.D.; Feihl, H.;
Science 294, 849-852, 2001.
D.: Jones, L.M.; Karet, U.
A:Authors: Kreft, J.; Kohn, M.; Kunst, F.; Kurapkut, G.; Madueno, E.; Maitournam, A.; Mat
ok, C.; Schluteler, T.; Simes, N.; Tietz, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend,
A:Title: Comparative genomics of *Listeria species*.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AH1387
A:Status: preliminary
A:Molecule type: DNA

A:Gene: SCOE94.19c

Query Match 9.2%; Score 153; DB 2; Length 228;

Best Local Similarity 37.7%; Pred. No. 0.00055;

Matches 46; Conservative 16; Mismatches 42; Indels 18; Gaps 6;

QY 219 VAGSTVNSGMW---FSGRD-----GDLINASNAGTVIQADHN-----MDGASIVYQHT 264

DB 101 LSNASFAONGGMAHKSIGDDFAVPFGITNNVVAHAGTIVKAGNGAGDGPAYGNAYIKHG 160

QY 265 NGPVSSYIHKDAQVKTGDTVRGQRIASMKV--QSSGAAL-FEFRISNR-GVYDPLTV 320

DB 161 NCTYSQYAHLSRINVIKGIQVKTGQSIASGNTGSSGPHLHFEIRTPYGSADVPAV 220

QY 321 LK 322

DB 221 LR 222

RESULT 29

AB2225 hypothetical protein alr3353 (imported) - Nostoc sp. (strain PCC 7120)

C:Species: Nostoc sp.

A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002

C:Accession: AB2225

R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Matanabe, A.; Iiziguchi, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.

DNA Res. 8, 205-213, 2001

A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena

A:Reference number: AB1807; MUID:21595285; PMID:11759840

A:Accession: AB2225

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-760 <KUR>

A:Cross-references: GB:BA000019; PIDN:BA075052.1; PID:g17132448; GSPDB:GN00179

A:Experimental source: strain PCC 7120

C:Genetics:

A:Gene: alr3353

Query Match 9.1%; Score 152; DB 2; Length 760;

Best Local Similarity 20.1%; Pred. No. 0.0031;

Matches 95; Conservative 41; Mismatches 112; Indels 224; Gaps 18;

QY 61 QYITD-SQGVNRYOKGDTVSKIAQRVGLMREIGHINNINSSTIYTGWLTLMSD 119

DB 294 QPLTETASATVTEKPEPDITLALASRYNTSVAEIVKVNLSNPQLKISQGLIIPS-- 351

QY 130 LKVRERSISSGVNTATTPS-----PVAVQSSRP--PVQGHFAVOK-- 157

DB 352 -----AIDSSSTIAQAPALISNRVQGTPTVPFVPTDLARVNSLSLVAQPPSIAND 405

QY 158 -----PTP-----PV-----VVKKP----- 168

DB 406 SVPITVPTPGADSETPVDITVPLESASAPAEIQGVGNVPIPKAFIEIORPQPGKRAAR 465

QY 169 -----TPPPVQO----- 177

DB 466 AKGDRLSIQAEIQRLOAKTRDOQGTNTVPPVANNNTALPIPVTSNNPAVTRPISR 525

QY 178 -----PAPVAPVTEAPFATGSSGVNQFR--YVQAT-----NPVRRFGTATV 219

DB 526 QOEIVAPVIAVPTPLPNDNHPVKQPFRAITLPPNEALNPEFLPNNAPQKPSVRVATPPV 585

QY 220 -----AGSTY-----TSNG 228

DB 586 RLNASBSLGRMGITVSPRLPPLAANDQYLPOAIDESVPPSPSTAFITPAKGVLTSGV 645

QY 229 MWFSGR-----DGLINASNAGTVIQADHNMDGAS--IVIOHTGPFVSYTHID 276

DB 646 GRRKGRMHRGIDIIAIPVGTVPALADGVVEKSGMNGGIGNTLVDIHHPDGSLTRVANSR 705

QY 277 AQVKTGDTVTRTGQRIASMKQPSGALF-----EFRIISNRGV-YVDPLTVL 321

DB 706 LTVQAGQVYRQGGQIHAM-----GSTGFTSGPHTHFEIHKTKGAINPIAML 752

RESULT 30

H86658 N-acetylglutamate-L-alanine amidase (EC 3.5.1.28) (imported) - Lactococcus lactis subsp.

C:Species: Lactococcus lactis subsp. lactis

C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001

C:Accession: H86658

R:Boletín, A.; Mincer, P.; Mager, S.; Jallón, O.; Malarme, K.; Weisenbach, J.; Ehrli

Genome Res. 11, 731-753, 2001

A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ssp

A:Reference number: A86625; MUID:21235186; PMID:11337471

A:Accession: H86658

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-439 <STO>

A:Cross-references: GB:AE005176; PID:g12723133; PIDN:AAK04370.1; GSPDB:GN00146

A:Experimental source: strain IL1403

C:Genetics:

A:Gene: acma

C:Keywords: hydrolase

Query Match 8.8%; Score 146.5; DB 2; Length 439;

Best Local Similarity 23.6%; Pred. No. 0.0038;

Matches 69; Conservative 39; Mismatches 107; Indels 77; Gaps 10;

QY 26 TTCIAGCASKPTYNST-----SGSGHRTSGSGGLAIGSQVITDSQGVN 72

DB 184 TALTKRYATDPVYGASLNRILISQVNLTRFDGASAGTSNSG-STATNNNSNTSST 242

QY 73 YQVKGDTVSKIAQRVGLMREIGHINNINSSTIYTGWLTLMSGDIAKREISISGVN 132

DB 243 YTVKSGDTLMGTSQKIGISVAQIQSANLKST-VITIGQLVLTSSSSNTNSSTSGN 301

QY 133 THTPSPVAVQSSRPVQGHFAVOKPPTPPVVKKPTPTPPVQGPAPVAPVT---EA 188

DB 302 SAGTTTPT-----STPKAPASQTIHKKVSGDTLWGLSV 336

QY 189 PFATGSSGVNQFRYPVQATNPVRRFGTAVAGSTVTSNGMWSGDDGLINASNAGTVI 248

DB 337 KYKTTIAQLKSNWHLNSDTIFIGQNLIVSQSAGSSSSSTG--SSSAST-- 382

QY 249 QADHNMDGASIVYQHTNGFVSSYIHKDAQVKTGDTV-----RTGQRIASMK 295

DB 383 SSTSNSSAAS-----NTSIH---KVVGDTLWGLSQKSGSPIASIK 420

RESULT 31

AG1147 P60 extracellular protein, invasion associated protein Iap (imported) - Listeria monocytogenes

C:Species: Listeria monocytogenes

C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001

C:Accession: AG1147

R:Glaser, P.; Frangoul, L.; Buchleser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker,

D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Duesuguet, O.; Ertian, K.D.; Feihl, H.

Science 294, 849-852, 2001

A:Authors: Krefel, J.; Kuhn, M.; Kunst, F.; Kurapkac, G.; Mañeno, E.; Maltournam, A.; Mat

ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,

A:Title: Comparative genomics of Listeria species

A:Reference number: AB1077; MUID:21537279; PMID:11679669

A:Accession: AG1147

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-482 <GLA>

A:Cross-references: GB:NC 003210; PIDN:CAC98661.1; PID:g16409958; GSPDB:GN00177

A:Experimental source: strain EGD-e

C:Genetics:

A:Gene: iap

Query Match 8.8%; Score 146.5; DB 2; Length 482;

C>Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999
 C/Accession: S76492
 R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
 O. K.; Okumura, S.; Shimoto, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
 DNA Res. 3, 109-116, 1996
 A/Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis*
 sp.
 A/Reference number: S74322; MUID:97061201; PMID:8905231
 A/Accession: S76492
 A/Status: nucleic acid sequence not shown; translation not shown
 A/Molecule type: DNA
 A/Residues: 1-715 <KAN>
 A/Cross-references: EMBL:D90915; GB:AB001339; NID:91653604; PID:BA18621.1; PID:d101935
 A/Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
 C/Genetics:
 A/Gene: nlpD

Query Match 8.3%; Score 138.5; DB 2; Length 715;
 Best Local Similarity 22.7%; Pred. No. 0.027;
 Matches 76; Conservative 44; Mismatches 128; Indels 87; Gaps 12;

QY YQVKGQDVTAKIQRVGLNWEIGHNNLSYITITQMTLWISGDL-----KYRERSI 127
 DB YQVKGQDVTAKIQRVGLNWEIGHNNLSYITITQMTLWISGDL-----KYRERSI 127
 QY 128 SSGVNTAHTPSPVAVQSSRPVQOH----- 152
 DB 438 PPSFSTGCTASTIYARAQAEPAQVQYQQLRAEVQNLQNTQPIRTPMVRSLATYARQVNN 497
 QY 153 ---PAVQKPTPPVVVKKPTPPPVVQGPAPVAPVTEAPAT--GSSGVN----- 198
 DB 498 EPVPMQARSPQALPAKFNQPRQDLAQORQYSPQARQSFSTVSGSQIYGAAPSPVQ 557
 QY 199 ---QRRYV-----GATNPVVR-----GATVAGSTVSN-----GMFSSGD- 235
 DB 558 YNDTITQIVQEVSELPGLSTPDPFRSPQGFNGITPAKGVFTISGFPFRMRHGRIDI 617
 QY 236 ---GDLINASAGTVIQADHNMDS--IYQHTNGVSSYIHKDAQVKTGDTVRTQ 289
 DB 618 AAPIGTPIMAASGCVVSGNNSGFGNLVKIRHDSVTYAHNNRLLVARGEVESQ 677
 QY 290 RIASN--KNQPSGAL--FEPRISRNGVYDPLTVL 321
 DB 678 QIAEWGSGRSTGPHLFEIRVGTNA--VNPVAL 711

RESULT 35
 C82736
 hypothetical protein XF0987 [imported] - Xylella faecidiosa (strain 9a5c)
 C/Species: Xylella faecidiosa
 C/Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
 C/Accession: C82736
 R:anonymous, The Xylella faecidiosa Consortium of the Organization for Nucleotide Sequen
 Nature 406, 151-157, 2000
 A/Title: The genome sequence of the plant pathogen *Xylella faecidiosa*.
 A/Reference number: A82515; MUID:20365717; PMID:10910347
 A/Note: for a complete list of authors see reference number A59328 below
 A/Accession: C82736
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-285 <SIM>
 A/Cross-references: GB:AB003937; GB:AE003849; NID:99105920; PID:AAF83797.1; GSPDB:GN001
 A/Experimental source: strain 9a5c
 R:Simpon, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
 Brites, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carter, H
 ae-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.
 submitted to GenBank, June 2000
 A/Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Prohm
 J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig
 chado, M.A.; Medeira, A.M.B.N.; Madeira, H.M.P.; Martino, C.L.; Marques, M.V.; Martins, E
 A.; Authors: Martins, E.M.P.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
 F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.
 Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak

A/Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
 M.; Teshako, M.H.; Valiada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
 A/Reference number: A59328
 A/Contents: annotation
 C/Genetics:
 A/Gene: XF0987

Query Match 8.2%; Score 136.5; DB 2; Length 285;
 Best Local Similarity 22.8%; Pred. No. 0.011;
 Matches 76; Conservative 49; Mismatches 126; Indels 83; Gaps 17;

QY 15 IKRLGIFGVITTCILAGC--ASKPTVNSTSGSGSHRTSGSGGLAISQVITDSQVPR 72
 DB 6 VNRL-LALGMLVITLPLPITAKKTKTOETRESEFPNVGALVIGT-----VPAG 56
 QY 73 YQVKGQDVTAKIQRVGLNWEIGH--INNLS-----SYITITQMTLWISGDL 119
 DB 57 SRVEYSGQVLRVDKGVVFGVGRNNITGPIILRQSEGVTVQVNTVTRQW----- 109
 QY 120 LKVERISISSGVNTAHTPSPVAVQSSRPVQOHAPVQKPTPPVVVKKPTTPPVQOPA 179
 DB 110 -----PLEVNVGVPKTVN-----PELTAAR-----ITRBA 137
 QY 180 PVADPTVTEAPATG--SSGVNQFRYPVQATNPVVRRTGATVAGSTVTSNGMFSGRD-- 235
 DB 138 ---QITDARQNRDNRSDFAQAFIWPV--QGRISGRFSARIYVGTATANG--HSGMDIA 189
 QY 236 ---GDLINASAGTVIQADHNMDS--IYQHTNGVSSYIHKDAQVKTGDTVRTQ 290
 DB 190 TATPTPKAPAPAGVITPAADPLVITGTVLIDHAGVSNFLHLSRIDVKGVDHDOGV 249
 QY 291 IASN--KNQPSGAL--FEPRISRNGVYDPLTVL 322
 DB 250 IGAVGATGRATGPHL--HMGWMPVVRIDPLVLE 282

RESULT 36
 D64158
 hypothetical protein H10756 - Haemophilus influenzae (strain Rd KW20)
 C/Species: Haemophilus influenzae
 C/Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 08-Oct-1999
 C/Accession: D64158
 R:Feilschmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A
 J.; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J
 D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghegan, N.S.M.
 Science 269, 496-512, 1995
 A/Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Frazer, C.M.; Smith, H.O.; Venter,
 A/Title: Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd.
 A/Reference number: A64000; MUID:95350630; PMID:7542800
 A/Accession: D64158
 A/Status: nucleic acid sequence not shown; translation not shown
 A/Molecule type: DNA
 A/Residues: 1-410 <TIGR>
 A/Cross-references: GB:U32760; GB:L42023; NID:91573764; PID:AA022415.1; PID:91573765; T
 A/Note: best homolog was a hypothetical protein from *Escherichia coli*

Query Match 8.1%; Score 135; DB 2; Length 410;
 Best Local Similarity 25.0%; Pred. No. 0.023;
 Matches 63; Conservative 36; Mismatches 111; Indels 42; Gaps 10;

QY 74 YQVKGQDVTAKIQRVGLNWEIGHNNLSYITITQMTLWISGDLKVERISISSGVNT 133
 DB 188 QKQQAALQKQAQEHQ-----STLNELNNKMLDQDKNT-----LKAEQALROEIOR 236
 QY 134 AHTPSPVAVQSSRPV--QOHAPVQKPTPPVVVKKPTTPPVVQOPAPVAPVTEADPAT 192
 DB 237 AEQAPARQEREREAALQROKAEKRT-----SKPYOPTVQEROLLN 278
 QY 193 GSSGV-----WQFRPVGATATPVVRRTGATVAGSTVTSNGMWSGRGDLINASAGTVI 248
 DB 279 STSLGAQAKQVSLPVSQS--ILHTFG--SIQAGEVWKKMVGASGTPKALAAARVI 334
 QY 249 QADH--NMDSAGSIYQHTNGVSSYIHKDAQVKTGDTVRTGORIASMKN--QPSGALFE 305

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 8, 2003, 10:50:52 ; Search time 22 Seconds

(without alignments)
607.063 Million cell updates/sec

Title: US-10-018-706-2

Perfect score: 1670

Sequence: 1 MTVTIAINSQKQKPIKRLGL.....LFEFRISHNGVYDPLVTLK 322

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	298	17.8	297	1 NLPD_PSEAE	P45882 pseudomonas
2	238	14.3	251	1 YGER_ECOLI	Q46798 escherichia
3	219.5	13.1	405	1 LPPB_HAETN	P44833 haemophilus
4	210	12.6	373	1 NLPD_SALTI	O56131.1 salmonella
5	209	12.5	377	1 NLPD_SALDI	P339700 salmonella
6	209	12.5	377	1 NLPD_SALDI	P40827 salmonella
7	202.5	12.1	379	1 NLPD_SALTI	P33648 escherichia
8	194	11.6	337	1 LPPB_HAESO	P36685 haemophilus
9	146.5	8.8	439	1 ACWA_LACLA	O9c14 lactococcus
10	146.5	8.8	484	1 P60_LISMO	P21171 listeria mo
11	146	8.7	524	1 P60_LISIV	O01837 listeria iv
12	144.5	8.7	523	1 P60_LISSE	O01838 listeria se
13	140.5	8.4	511	1 P60_LISGR	O01835 listeria gr
14	139.5	8.4	467	1 P60_LISIN	O01836 listeria in
15	137.5	8.2	524	1 P60_LISWE	O01839 listeria we
16	135	8.1	410	1 YIRB_HAETN	P44864 haemophilus
17	132	7.9	475	1 YERB_HAETN	P44693 haemophilus
18	130.5	7.8	437	1 ACWA_LACUC	Q46803 lactococcus
19	126.5	7.6	374	1 PRUB_ACHLY	P27458 actinobact
20	124	7.4	1742	1 GUNA_CALSA	P22534 caldocellum
21	123.5	7.4	666	1 MUR2_ENTHR	P23946 enterococc
22	118	7.1	633	1 LAI7_YEAST	Q12446 saccharomy
23	118	7.1	1192	1 RTM4_HUMAN	O9nec3 homo sapien
24	116.5	7.0	282	1 HES1_MOUSE	P35428 mus musculu
25	116	6.9	346	1 NLPD_BUCAL	P57493 buchnera ap
26	116	6.9	334	1 PRP1_LYCES	Q00451 lycopersico
27	115	6.9	97	1 NLPD_YEREN	P47764 yerisida en
28	114.5	6.9	2004	1 MOZ_HUMAN	O92794 homo sapien
29	114	6.8	281	1 HES1_RAT	O04666 rattus norv
30	113.5	6.8	1331	1 MAMB_CALSA	P22333 caldocellum
31	109.5	6.6	280	1 HES1_HUMAN	Q14469 homo sapien
32	109.5	6.6	1181	1 MYSC_DICDI	P42522 dictyosteli
33	109	6.5	907	1 VGP3_EBV	P03200 Epstein-bar

34	108.5	6.5	283	1 EXTN_SORBI	P24152 sorghum bic
35	108.5	6.5	555	1 GPI_CHLRE	Q94P66 chlamydomon
36	108.5	6.5	3591	1 FRAB_BORPE	P12255 bordetella
37	108	6.5	1593	1 AT12_HUMAN	P58397 homo sapien
38	107.5	6.4	671	1 ALYS_ENTFA	P37710 enterococcu
39	107.5	6.4	777	1 PRZE_MYXXA	P18769 myxococcu
40	107.5	6.4	2167	1 SHK1_RAT	Q9wv48 rattus norv
41	107	6.4	553	1 ODP2_ALCEU	P35829 lactobacilli
42	107	6.4	553	1 ODP2_ALCEU	O59098 alcaligenes
43	106.5	6.4	3176	1 CA36_HUMAN	P12111 homo sapien
44	106	6.3	532	1 SPG7_DICDI	P22698 dictyosteli
45	106	6.3	672	1 HMW3_MYCPN	Q50360 mycoplasma

ALIGNMENTS

RESULT 1
ID NLPD_PSEAE STANDARD, PRT, 297 AA.
AC P45682:
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Lipoprotein nlpd/lppb homolog precursor.
GN P43623.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PAO1;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
Hickey M.J., Brinkman F.S.L., Huftagle W.O., Kowalik D.J., Lagrou M.,
Gardner R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
Brody L.L., Coulter S.N., Folger K.R., Kas A., Laidig K., Lim R.M.,
Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.,
"Complete genome sequence of Pseudomonas aeruginosa PAO1, an
opportunistic pathogen."
RT Nature 406:959-964(2000).
RL -1- SUBCELLULAR LOCATION: Attached to the inner membrane by a lipid
anchor (potential).
CC -1- SIMILARITY: BELONGS TO THE E. COLI NLPD / HAEMOPHILUS LPPB FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D26134; BA05130.1; -
CC EMBL: AE004782; AA07011.1; -
CC MEROPS: M37. UPW; -
CC InterPro: IPR002482; LysM.
CC InterPro: IPR002886; Peptidase_M37.
CC InterPro: IPR000437; Prok_lipoprot.
CC Pfam: PF01476; LysM; 1.
CC Pfam: PF01551; Peptidase_M37; 1.
CC SMART: SMO0257; LysM; 1.
CC PROSITE: PS00013; PROKAR_LIPOPROTEIN; FALSE_NEG.

Inner membrane; Lipoprotein; Signal; Complete proteome.

Query Match
Best Local Similarity 17.8%; Score 298; DB 1; Length 297;
Matches 91; Conservative 50; Mismatches 106; Indels 78; Gaps 12;

20 LIFGVITTCILAGCASKP-----TYNSTSGSGHRTSGSGGLAIGSOVITDSQGVNKR 72
25 LLGAVVCSILAACSSPPGCVKVDNNSAPAAARRTPVTSQ-----Q 68
73 YVKGQDVTYSKIAQRYGLMWRREIGHINNNSSYTIYTGWLTLMSGDLKVRERSISSGVN 132
69 YVRRGDTLYSIAFRFGMDMKALARNGLAPPTIIVGQ----- 107
133 TAHPSPVAVQ--SSRPVQOHPAVOKPTPPVV--VVKKPTPTPPVVQ-----QPAVAPP 184
108 -----AIQFGFRASTQ--PSVAKNTPVAVAPVAKTPTPPVPAVSTSVPAKAP-ADA 156
185 VTEAPPATSSGVVQ-----FRYPGATNPVVRFRGTATVAGSTVTSNGMFPSGRDDL 238
157 STTPPSGATPVVAVGAVGAMVAPASGT--LIGRF-----ASNGSLNKIDIDAGLQGP 209
239 INSNNGVTYVQADHNDGAS--IVIOHTNGFVSYYHIKDAQYKGTGTVRTGQRIASMGK 296
210 VLAASGTVVYAGSGLRGVELVIKHNETVYVSAIGHNRLLVREGQVVKVGSIAEMQS 269
297 QPSGALFERRISRGVYVDELTVL 321
270 TGTDRVKTFRIRKQKPPVLPQYL 294

RESULT 2
YGER_ECOLI STANDARD; PRT; 251 AA.
ID YGER_ECOLI STANDARD; PRT; 251 AA.
AC Q46798; 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE Hypothetical lipoprotein Yger precursor.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OC NCBI_Taxid=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; Pubmed=9278503;
RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Colado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RT Science 377:1453-1474(1997);
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
CC (Potential).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U28375; AAA83046.1; ALT INIT.
CC EMBL; AEO00369; AAC75903.1; ALT INIT.
CC EcoGene; EG13048; Yger.
CC InterPro; IPR002482; LysM.

InterPro; IPR002886; Peptidase_M37.
DR Pfam; PF01476; LysM; 1.
DR Pfam; PF01551; Peptidase_M37; 1.
DR SMART; SMO0257; LysM; 1.
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW Hypothetical protein; Membrane; Lipoprotein; Signal;
KW Complete proteome.
FT SIGNAL 1 25
FT CHAIN 26 251
FT LIPID 26 251
SQ SEQUENCE 251 AA; 26564 MW; FCD6235D9790475 CRC64;

Query Match
Best Local Similarity 14.3%; Score 238; DB 1; Length 251;
Matches 78; Conservative 46; Mismatches 103; Indels 86; Gaps 8;

16 KRLGIIFGVITTCILAGCASKPTNYSYSGSGHRTSGSGGLAIGSOVITDSQGVNRYOV 75
9 KSLGIIVMLISVGLLAGC-----SGSKSDPTGYSG-----SVYTV 44
76 KQDVTYSKIAQRYGLMWRREIGHINNNSSYTIYTGWLTLMGDLK---VRERSISSGV 131
45 KRGDLYRISRTTGSVVELRLNGISPPYIEVGOKLKL--GGAKSSITRKSTAKSTT 102
132 NTAA--TPSPVAVQSSRPVQOHPAVOKPTPPVVVVKKPTPTPPVVOQPAVAPPVTEAPF 190
103 KTA SVTPSSAVPKSSWPPVQRCWLMPTTGKVM----- 138
191 ATGSSGVNQFRYPVATNPVVRFRGTATVAGSTVTSNGMFPSGRDDLINSNAGTYVQA 250
139 STADGG-----NKGIDISARGTPIVYAGAGKVVV 169
251 DHNMDSAS--IVIOHTNGFVSYYHIKDAQYKGTGTVRTGQRIASMKNOGSAALFERRI 308
170 GNGLRGVGNLIMIKHSEDIYITAVANDTLMVNNQSVYAGGKXIAIMGSTDPAVRLHQI 229
309 SRNGVYVDELTVL 321
230 RYRATADLPRLYL 242

RESULT 3
LPPB_HAEIN STANDARD; PRT; 405 AA.
ID LPPB_HAEIN STANDARD; PRT; 405 AA.
AC P44833; 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Outer membrane antigenic lipoprotein B precursor.
GN LPPB OR H10706.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OC NCBI_Taxid=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / ATCC 51907;
RX MEDLINE=95350630; Pubmed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shiley R., Liu L.-I., Giordano A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Usterback T.R., Hanna M.C., Nguyen T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.U., Fritchman J.U., Geoghegan N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Frazer C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd.";
RT Science 269:496-512(1995).
CC -1- FUNCTION: MAY BE A VIRULENCE DETERMINANT.
CC -1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
CC anchor.

CC -1- SIMILARITY: BELONGS TO THE E. COLI NLPD / HAEMOPHILUS LPPB FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U32753; AAC2363.1; -
 CC MEROPS: M37.UPM; -
 CC TIGR: H10706; -
 CC InterPro: IPR002482; LysM.
 CC InterPro: IPR002886; Peptidase_M37.
 CC Pfam: PF01476; LysM; 1.
 CC Pfam: PF01551; Peptidase_M37; 1.
 CC SMART: SM00257; LysM; 1.
 CC PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
 CC Outer membrane; Lipoprotein; Signal; Antigen; Virulence;
 CC Complete proteome.
 CC SIGNAL 1 17 PROBABLE.
 CC CHAIN 18 405 OUTER MEMBRANE ANTIGENIC LIPOPROTEIN B.
 CC LIPID 18 18 N-ACYL DIGLYCERIDE (PROBABLE).
 CC SEQUENCE 405 AA; 42505 MW; 4134FB6B93EF57F CRC64;
 SQ
 Query Match 13.1%; Score 219.5; DB 1; Length 405;
 Best Local Similarity 27.8%; Pred. No. 3e-08;
 Matches 76; Conservative 46; Mismatches 114; Indels 37; Gaps 10;
 Oy 71 NRYQYQGDIVTSKIAQRYGLNREIGHININSSYTIYQGLTMSGDLKVRERSISG 130
 Db 145 NRYQYQGDIVTSKIAQRYGLNREIGHININSSYTIYQGLTMSGDLKVRERSISG 130
 Oy 131 VNTAHTSPVAVQSRPPYQCHPANKPTP-----PVVVKKPTPTPPV 175
 Db 203 -----VKQPAVATVATPTVK--PAVTV--TPGANGTQIGSDGITIGPKSEAGTSPSVEVA 254
 Oy 176 QQPAPVAPVTEAPATGSSGVWQFRYPVVGATNPVRRPGRTA--TVAGSTV---SNGMW 230
 Db 255 TSSIVTSSVANNANSTPINSNVA--PI-ASN-VVWQMPISGNIIOGSSITDGNKIID 309
 Oy 231 FSGRDGDLINMSNACTVIOADHNDGAS--IVIOHTNGFVSYSYIHKDAQVKTGPTVETG 288
 Db 310 ISGSGQAVKAAARIVYAGNALMGYNLLITIKNDPFLSAYANDKILVADQGEVAG 369
 Oy 289 QRISMKQPSGALFERRISNGYVDPPLTVL 321
 Db 370 QDIAMGSSGTNTVTLAFERIKRSVDPVRYL 402
 RESULT 4
 NLPD_SALTI STANDARD; PRT; 373 AA.
 ID NLPD_SALTI
 AC Q56131;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Lipoprotein nlpd precursor.
 GN NLPD OR STY3050.
 OS Salmonella typhi.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Salmonella.
 CX NCBI_TaxId=601;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CT18;
 RX MEDLINE=21534947; PubMed=11677608;
 RA Parthill J., Dougan G., James K.D., Thomson N.R., Pickard D., Main J.,
 RA Churcher C., Mungall K.L., Bentley S.D., Holden M.R., Sebahia M.,
 RA Baker S., Bauman D., Brooks K., Chillingworth T., Connor P.,
 RA Cronin A., Davis P., Davies R.M., Hens N., White N., Farrar J.,
 RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagers K.,

RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
 RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
 RA Whitehead S., Barrall B.G.;
 RT "Complete genome sequence of a multiple drug resistant Salmonella
 RT enterica serovar Typhi CT18".
 RL Nature 413:848-852(2001).
 RN [2]
 RP SEQUENCE OF 255-373 FROM N.A.
 RC STRAIN=Ty2a;
 RX MEDLINE=95220644; PubMed=7705608;
 RA Robbe-Saule V., Coyanult C., Norel F.;
 RT "The live oral typhoid vaccine Ty2a is a rpos mutant and is
 RT susceptible to various environmental stresses."
 RL FEMS Microbiol. Lett. 126:171-176(1995).
 RN [3]
 RP SEQUENCE OF 269-373 FROM N.A.
 RC STRAIN=Ty2;
 RX MEDLINE=99118315; PubMed=9919662;
 RA Robbe-Saule V., Norel F.;
 RT "The rpos mutant allele of Salmonella typhi Ty2 is identical to that
 RT of the live typhoid vaccine Ty2a".
 RL FEMS Microbiol. Lett. 170:141-143(1999).
 CC -1- FUNCTION: MAY BE INVOLVED IN STATONARY-PHASE SURVIVAL.
 CC -1- SUBCELLULAR LOCATION: Attached to the inner membrane by a lipid
 CC anchor (Potential).
 CC -1- SIMILARITY: BELONGS TO THE E. COLI NLPD / HAEMOPHILUS LPPB FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AL627276; CAD06031.1; -
 CC EMBL: X81641; CA57297.1; -
 CC EMBL: Y17610; CA476806.1; -
 CC MEROPS: M37.UPM; -
 CC InterPro: IPR002482; LysM.
 CC InterPro: IPR002886; Peptidase_M37.
 CC Pfam: PF01476; LysM; 1.
 CC Pfam: PF01551; Peptidase_M37; 2.
 CC SMART: SM00257; LysM; 1.
 CC PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
 CC Inner membrane; Lipoprotein; Repeat; Signal; Complete proteome.
 CC SIGNAL 1 25
 CC CHAIN 26 373 LIPOPROTEIN NLPD.
 CC LIPID 26 26 N-ACYL DIGLYCERIDE (BY SIMILARITY).
 CC DOMAIN 199 246 4 X 7 AA APPROXIMATE REPEATS.
 CC REPEAT 199 205 1.
 CC REPEAT 221 227 2.
 CC REPEAT 233 239 3.
 CC REPEAT 240 246 4.
 CC SEQUENCE 373 AA; 39183 MW; DD075BB44098FF7 CRC64;
 SQ
 Query Match 12.6%; Score 210; DB 1; Length 373;
 Best Local Similarity 23.2%; Pred. No. 1.2e-07;
 Matches 84; Conservative 56; Mismatches 138; Indels 84; Gaps 13;
 Oy 30 LAGCASK-----PYNSTSGSGSHRTSG-----SGGL-----AIGS 60
 Db 23 LAGTSSSNPPAPVTSVDSGSSSTNGMLITPPKMGATPQQAPOIQPVPTQGMQT 82
 Oy 61 QVITDSQ-----GVF-----NRYQYQGDIVTSKIAQRYGLNREIGHI 98
 Db 83 QPVTQEPQVOMENGRIVNRYQYGNIPKSGSYTGSTYVTKGDTLLYIAMITGNDPRDAQR 142
 Oy 99 NNINSSYTIYQGL-----TLMSGDLKVRERSISGVNTAHTPSPAVAVQSRPPV- 149
 Db 143 NSIAPFSLNVGQLTVQGNASGPIPTGNAITQDAQAQGVVTKSAQNSTVAVASQPTT 202
 Oy 150 -----QOHPAVQKPT--PVVVVKKPTPTPPVQCPAPVAPVTEAPATGSSGVWQFR 201

```

DB      203 YSESGEGSANKM|LNKKPAGTVVAPVTA|PTVSTTEPNASSTISAP|-----ISAMR 255
QY      202 YPVGATN|VYRRFRGTATVAGSTVTSNGMWFSGRGGDL|INASVAGTVIOADHNM|DGA--I 259
DB      256 WP--|TDGKVIENFG-ASEEG|-----NK|ID|D|GSKGQALVATDGRVYVAGNALRGV|NLI 308
QY      260 VI|GHTNGVSSYTHIKDAQVKTG|D|VTRGQRIASMKQPSG|AALFEPRISRN|GVYVDELT 319
DB      309 I|K|N|D|D|Y|L|S|A|Y|A|N|D|T|M|L|V|R|E|Q|E|V|K|A|G|K|I|A|T|G|S|T|G|S|T|R|H|F|E|I|Y|K|G|S|V|N|P|R 368
QY      320 VL 321
DB      369 YL 370

RESULT 5
NLPD_SALDU STANDARD; PRT; 377 AA.
AC P39700;
DT 01-FEB-1995 (Rel. 31, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Lipoprotein nlpD precursor.
GN NLPD.
OS Salmonella dublin.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_Taxid=98360;
RN [1]
RP SEQUENCE FROM N.A., AND INDUCTION.
RC STRAIN=lane;
RX MEDLINE=9938748; PubMed=973354;
RA Paesold G., Krause M.;
RT "Analysis of rpos mRNA in Salmonella dublin: identification of
RT multiple transcripts with growth-phase dependent variation in
RT transcript stability."
RT J. Bacteriol. 181:1264-1268(1999).
RN [2]
RP SEQUENCE OF 284-377 FROM N.A.
RC STRAIN=lane;
RA Krause M.W., El-Gedaily A.;
RL Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: MAY BE INVOLVED IN STATIONARY-PHASE SURVIVAL.
CC -1- SUBCELLULAR LOCATION: Attached to the inner membrane by a lipid
CC anchor (Potential).
CC -1- SIMILARITY: BELONGS TO THE E. COLI NLPD / HAEMOPHILUS LPPB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AJ006131; CA06881.1; -
DR EMBL: X82129; CA57639.1; -
DR MEROPS: M37.0PW; -
DR InterPro: IPR002483; LysM.
DR InterPro: IPR002886; Peptidase_M37.
DR InterPro: IPR000437; Prok_Lipoprot.
DR Pfam: PF01476; LysM; 1.
DR SMART: SM00257; LysM; 1.
DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
DR Inner membrane; Lipoprotein; Repeat; Signal.
FT CHAIN 1 25
FT SIGNAL 1 25
FT LIPID 26 377
FT DOMAIN 203 250
FT REPEAT 203 209
FT REPEAT 225 231

```

```

FT REPEAT 237 243 3.
FT REPEAT 244 250 4.
SQ SEQUENCE 377 AA; 39671 MW; DC7F59B943F86512 CRC64;
Query Match 12.5%; Score 209; DB 1; Length 377;
Best Local Similarity 23.0%; Pred. No. 1.5e-07;
Matches 84; Conservative 56; Mismatches 138; Indels 88; Gaps 13;

QY 30 LAGGASK-----PYNSTSGSGSHRTSG-----SGGLA----- 57
DB 23 LAGCTSSNPAP|P|T|V|S|D|S|S|S|V|T|N|S|G|L|I|T|P|P|K|G|A|T|Q|T|P|Q|A|P|Q|I|P|V|Q|R|P|V|T|Q 82
QY 58 -IGSOVITDSQ-----GVP-----NRVOYKQD|TVASK|A|O|Y|G|L|N|W|R|E 94
DB 83 PMQTPVTEQPVCMENGRIVYNNROYNI|PKGSYTGSG|TYVKKG|DLFY|A|W|I|G|N|D|F|R|D 142
QY 95 IGHINN|LNS|TYTTTG|G|L|-----TLMSGLK|V|E|R|S|S|G|V|N|T|A|H|P|S|V|A|V|Q|S|R 146
DB 143 LAQRN|S|I|S|A|P|Y|S|I|N|V|G|T|L|O|V|G|N|A|S|G|P|T|G|N|A|I|T|Q|D|A|Q|G|V|T|R|S|A|N|S|I|V|A|A|S|Q 202
QY 147 PPV-----QHPAVQKPT--PPVVVYK|E|P|T|P|P|V|Q|P|A|V|P|V|T|A|P|F|A|T|G|S|G|V 197
DB 203 PTIYVSEGSQ|S|A|N|K|M|L|P|N|K|P|A|G|T|V|V|A|P|V|T|A|P|V|T|E|P|N|A|S|T|S|T|S|A|P|-----I 255
QY 198 MQFRYPV|G|A|T|N|P|V|Y|R|R|G|T|A|T|V|A|G|S|T|V|T|S|N|G|M|F|S|G|R|G|D|L|I|N|S|N|A|G|T|V|I|O|A|D|H|N|D|G|A 257
DB 256 SAMRWP--TDGKVIENFG-ASEEG|-----NK|ID|D|GSKGQALVATDGRVYVAGNALRGV 308
QY 258 S--IV|GHTNGVSSYTHIKDAQVKTG|D|VTRGQRIASMKQPSG|AALFEPRISRN|GVY 315
DB 309 G|N|L|I|K|N|D|D|Y|L|S|A|Y|A|N|D|T|M|L|V|R|E|Q|E|V|K|A|G|K|I|A|T|G|S|T|G|S|T|R|H|F|E|I|Y|K|G|S|V 368
QY 316 DPLTVL 321
DB 369 NPLRYL 374

RESULT 6
NLPD_SALTY STANDARD; PRT; 377 AA.
AC P40827;
DT 01-FEB-1995 (Rel. 31, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Lipoprotein nlpD precursor.
GN NLPD OR STM2925.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_Taxid=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=2153448; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du P., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2."
RT Nature 413:852-856(2001).
RN [2]
RP SEQUENCE OF 280-377 FROM N.A.
RC STRAIN=ATCC 14028;
RX MEDLINE=9436857; PubMed=808645;
RA Prince R.W., Fang F.C., Libby S.J.;
RT Cloning and sequencing of the gene encoding the rpos (KatF) sigma
RT factor from Salmonella typhimurium 14028s."
RN Biochim. Biophys. Acta 1219:196-200(1994).
RN [3]
RP SEQUENCE OF 298-377 FROM N.A.
RC STRAIN=C52;

```

```

RX MEDLINE=95050252; PubMed=7961444;
RA Kowarz L., Coyanalt C., Robbe-Saule V., Norel F.;
RT "The Salmonella typhimurium katF (rpos) gene: cloning, nucleotide
RT sequence, and regulation of spvR and spvABCD virulence plasmid
RT genes.";
RL J. Bacteriol. 176:6852-6860(1994).
CC -1- FUNCTION: MAY BE INVOLVED IN STATIONARY-PHASE SURVIVAL.
CC -1- SUBCELLULAR LOCATION: Attached to the inner membrane by a lipid
CC anchor (Potential).
CC -1- SIMILARITY: BELONGS TO THE E. COLI NLPD / HAEMOPHILUS LPPB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE008833; AAL21805.1; -
DR EMBL; U05011; -; NOT_ANNOTATED_CDS.
DR EMBL; X77752; -; NOT_ANNOTATED_CDS.
DR MEROPS; M37.UPM; -
DR Skygene; SG10513; nlpD.
DR InterPro; IPR002886; Peptidase_M37.
DR InterPro; IPR000437; Prok_lipoprot.
DR Pfam; PF01551; Peptidase_M37; 1.
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KM Inner membrane; lipoprotein; Repeat; Signal; Complete proteome.
FT SIGNAL 1 25
FT CHAIN 1 25
FT LIPID 26 377
FT DOMAIN 203 250 4 X 7 AA APPROXIMATE REPEATS.
FT REPEAT 203 209 1.
FT REPEAT 225 231 2.
FT REPEAT 237 243 3.
FT REPEAT 244 250 4.
FT CONFLICT 377 377 R -> P (IN REF. 2).
SQ SEQUENCE 377 AA; 39641 MW; 27C29D77A145BFO CQC64;

Query Match 12.5%; Score 209; DB 1; Length 377;
Best Local Similarity 23.0%; Pred. No. 1.5e-07;
Matches 84; Conservative 56; Mismatches 138; Indels 88; Gaps 13;

QY 30 LAGCASK-----PTNSTSGSGSHRTSG-----SGGLA----- 57
DB 23 LAGCTSSNPAPATVTSVDSGSSSTNTNSMLITPPPKMGATTTQCTPQAPQIQPVQPYTQ 82
QY 58 -IGSQVITDSQ-----GVP-----NRYOVKQDPTVSKIAQRYGLNMR 94
DB 83 PMQGPVTEQPVQMGNGRIIVNRQYGNIPKSGSYTGSGTYTKGDTLRYIAMITGDNFRD 142
QY 95 IGHINNLSYTYTGQWL-----TLMSGDLKVRERSISSGVNTAHTBSPVAVGSSR 146
DB 143 LAQRNISIAPYSINAGTLOVGNASGTPITGNAITQDAQAQGVTVTSQAQNSTVAVASQ 202
QY 147 PPV-----QGHRAVQKPT--PPVVVVKKPTPPPVVQGPAPVAPVYTAEPFAAGSSGV 197
DB 203 PTTYSSESGEESANKMLPNKPKACTVATVATVATVSTTEBNASTSTSA-----I 255
QY 198 MGFRTVGAATNPVVRFGTATVAGSTVTSNGMFGSGRGDGLINASNAAGTVIQADHNMDGA 257
DB 256 SAMWRP--TDGVINFG--ASREG--NKGDIDAGSGGAIVATADRVRVYAGNALGX 308
QY 258 S--IYIQTNGFVSSYHIKQAQVKTGTTGRTGQRIASMKNOPSGAALFEPRISNGYV 315
DB 309 GNLIIKINDVYLSAYAHNDTWLVREQOEKVGAKGIATAGSTGTSTLHEIRKXGKSV 368
QY 316 DPLTVL 321
DB 369 NPLRYL 374

```

```

RESULT 7
NLPD_ECOLI STANDARD; PRT; 379 AA.
ID P33648;
AC P33648;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Lipoprotein nlpD precursor.
GN NLPD OR B2742.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MP180.
RX MEDLINE=94179096; PubMed=8132457;
RA Ichikawa J.K., Li C., Fu J.C., Clarke S.;
RT "A gene at 59 minutes on the Escherichia coli chromosome encodes a
RT lipoprotein with unusual amino acid repeat sequences.";
RL J. Bacteriol. 176:1630-1638(1994).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=9742617; PubMed=9278503;
RA Blatter F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
[3]
RP SEQUENCE OF 99-379 FROM N.A.
RC STRAIN=K12 / DH1.
RX MEDLINE=94268497; PubMed=8208244;
RA Takahashi Y., Tanaka K., Takahashi H.;
RT "Structure of the 5' upstream region and the regulation of the rpos
RT gene of Escherichia coli.";
RL Mol. Gen. Genet. 243:525-531(1994).
CC -1- FUNCTION: MAY BE INVOLVED IN STATIONARY-PHASE SURVIVAL.
CC -1- SUBCELLULAR LOCATION: Attached to the inner membrane by a lipid
CC anchor (Potential).
CC -1- SIMILARITY: BELONGS TO THE E. COLI NLPD / HAEMOPHILUS LPPB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L07869; AAA17875.1; -
DR EMBL; U29579; AAA69252.1; -
DR EMBL; AE000358; AAC75784.1; -
DR EMBL; D17549; BAA04487.1; -
DR MEROPS; M37.UPM; -
DR SWISS-2DPAGE; P33648; COLI.
DR EcGene; EG12111; nlpD.
DR InterPro; IPR002482; LysM.
DR InterPro; IPR002886; Peptidase_M37.
DR Pfam; PF01476; LysM; 1.
DR Pfam; PF01551; Peptidase_M37; 1.
DR SMART; SM00257; LysM; 1.
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KM Inner membrane; lipoprotein; Repeat; Signal; Complete proteome.
FT SIGNAL 1 25
FT CHAIN 1 25
FT LIPID 26 379
FT DOMAIN 66 97
FT REPEAT 66 73
FT REPEAT 74 81

```

FT	REPEAT	82	89	1-3	1-4	(APPROXIMATE).
FT	REPEAT	90	97	252	4 X 7 AA	APPROXIMATE REPEATS.
FT	DOMAIN	205	252	211	2-1.	
FT	REPEAT	205	211	227	2-3	
FT	REPEAT	227	233	2-2.		
FT	REPEAT	239	245	2-3		
FT	REPEAT	246	252	2-4.		
FT	CONFLICT	139	139	G -> A (IN REF. 3).		
FT	SEQUENCE	379 AA;	40149 MW;	ABE6A2B8456105FE CRC64;		
Query Match	Similarity	12.1%;	Score 20.5;	DB: 1;	Length 379;	
Best Local	72;	Conservative	42;	Mismatches 113;	Indels 49;	Gaps 9;
OY	73	YQVKGQDTVSKIAQRYGLNMFREIGHINNLLNSYITTYGQMLTMSGDLKRYERSSISGVN	132			
Db	123	YTVKKGDTLFYIAMITGNDPFDLQARNNIQAPYALNWGO--TLQVG-----	167			
OY	133	TAHTPSPVAVOSSRPVQOHPYAKP--TPPVVYKKPT-----	P 170			
Db	168	ASGTPITGNAITQADAAEQCVIKFPAONSTVVAASOPTITYSBSSGEOANMKLPNNKP	227			
OY	171	TPPVQOQAPAPAPPVTEAPFATG--SSGVMQFRYPVATNPVYRRFGTAVAGSTVTSN	227			
Db	228	TATVTAATVATPASTTEPTVTSSTSTPSTWMP--TEGKVIETFG-ASEGG----	NK 280			
OY	228	GMWFSGRGDDILINSMAGTVIQADHNMGAS--IVIQHTNGFVSYSIHKDAQVKTGDTV	285			
Db	281	GIDLAGSGQAIIATADRGVYVAGNALGYNLLIIRKNDYLSAAYANDTMLVREQEV	340			
OY	286	RTGQRIASMKQNPQSGALFEPRISNGVYVPLTVL	321			
Db	341	KAGQKIATMGSTGTSSTRLHFEIRKXGSMVPLTVL	376			
RESULT 8						
LPBP_HABSO	ID	LPBP_HABSO	STANDARD:	PRT;	337 AA.	
AC	P36685;					
DT	01-JUN-1994 (Rel. 29, Created)					
DT	01-NOV-1995 (Rel. 32, Last sequence update)					
DT	01-OCT-1996 (Rel. 34, Last annotation update)					
DE	Outer membrane antigenic lipoprotein B precursor (Fragment).					
GN	LPBP.					
OS	Haemophilus somnus.					
OC	Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;					
OC	Haemophilus.					
OX	NCBI_TaxID=731;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN=HS25;					
RC	MEDLINE=93239280; PubMed=8478068;					
RA	Theisen M., Rioux C.R., Potter A.A.;					
RT	"Molecular cloning, nucleotide sequence, and characterization of					
RT	lpbB, encoding an antigenic 40-kilodalton lipoprotein of Haemophilus					
RT	somnus.";					
RU	Infect. Immun. 61:1793-1796(1993).					
RU	[2]					
RP	CONCEPTUAL TRANSLATION.					
RA	Baitoch A.;					
RU	Unpublished observations (Aug-1995).					
CC	-1- FUNCTION: MAY BE A VIRULENCE DETERMINANT.					
CC	-1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid					
CC	anchor.					
CC	-1- SIMILARITY: BELONGS TO THE E. COLI NUPD / HAEMOPHILUS LPBP FAMILY.					
CC	-1- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; THREE FRAMESHIFTS HAD					
CC	TO BE INTRODUCED IN POSITIONS 264, 266 AND 328.					
CC	-----					
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration					
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -					
CC	the European Bioinformatics Institute. There are no restrictions on					
CC	use by non-profit institutions as long as its content is in no way					
CC	modified and this statement is not removed. Usage by and for commercial					

CC	entities requires a license@sb-rib.ch).	CC	entities requires a license@sb-rib.ch).
CC	cc send an email to license@sb-rib.ch .	CC	cc send an email to license@sb-rib.ch .
DR	EMBL; L10653; AAA72348.1; ALT_FRAME.	DR	EMBL; L10653; AAA72348.1; ALT_FRAME.
DR	MEROPS; M37.UPW; -.	DR	MEROPS; M37.UPW; -.
DR	InterPro; IPRO02482; LysM.	DR	InterPro; IPRO02482; LysM.
DR	InterPro; IPRO02886; Peptidase_M37.	DR	InterPro; IPRO02886; Peptidase_M37.
DR	Pfam; PF01476; LysM; 1.	DR	Pfam; PF01476; LysM; 1.
DR	Pfam; PF01551; Peptidase_M37; 1.	DR	Pfam; PF01551; Peptidase_M37; 1.
DR	SMART; SMO0257; LysM; 1.	DR	SMART; SMO0257; LysM; 1.
DR	PROSITE; PS00013; PROKAR LIPOPROTEIN; 1.	DR	PROSITE; PS00013; PROKAR LIPOPROTEIN; 1.
DR	Outer membrane; Lipoprotein; Signal; Antigen; Virulence.	DR	Outer membrane; Lipoprotein; Signal; Antigen; Virulence.
FT	SIGNAL	FT	SIGNAL
FT	CHAIN	FT	CHAIN
FT	LIPID	FT	LIPID
FT	NON_TER	FT	NON_TER
SO	SEQUENCE	SO	SEQUENCE
Query Match	Best Local Similarity	Query Match	Best Local Similarity
Matches	61; Conservative	Matches	61; Conservative
QY	73	QY	73
Db	120	Db	120
QY	130	QY	130
Db	180	Db	180
QY	190	QY	190
Db	230	Db	230
QY	250	QY	250
Db	269	Db	269
QY	308	QY	308
Db	329	Db	329
RESULT 9		RESULT 9	
ACMA_LACLA	STANDARD;	ACMA_LACLA	STANDARD;
ID	Q9C1T4;	ID	Q9C1T4;
AC	16-OCT-2001 (Rel. 40, Created)	AC	16-OCT-2001 (Rel. 40, Created)
DT	16-OCT-2001 (Rel. 40, Last sequence update)	DT	16-OCT-2001 (Rel. 40, Last sequence update)
DT	15-JUN-2002 (Rel. 41, Last annotation update)	DT	15-JUN-2002 (Rel. 41, Last annotation update)
DE	Probable N-acetylmuramidase precursor (BC 3.2.1.17) (Peptidoglycan hydrolyase) (Autolysin) (Lysosome).	DE	Probable N-acetylmuramidase precursor (BC 3.2.1.17) (Peptidoglycan hydrolyase) (Autolysin) (Lysosome).
OS	ACMA OR L10272.	OS	ACMA OR L10272.
OS	Lactococcus lactis (subsp. lactis) (Streptococcus lactis).	OS	Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.	OC	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX	NCBI_TaxID=1360;	OX	NCBI_TaxID=1360;
RN	SEQUENCE FROM N.A.	RN	SEQUENCE FROM N.A.
RP	STRAIN=IL1403;	RP	STRAIN=IL1403;
RC	MEDLINE=2123186; PubMed=1137471;	RC	MEDLINE=2123186; PubMed=1137471;
FX	Bojoutin A., Wincker P., Manger S., Jallion O., Malarme K.,	FX	Bojoutin A., Wincker P., Manger S., Jallion O., Malarme K.,
RA	Weissenbach T., Ehrlich S.D., Sorokin A.;	RA	Weissenbach T., Ehrlich S.D., Sorokin A.;
RT	"The complete genome sequence of the lactic acid bacterium Lactococcus lactis ssp. lactis IL1403."	RT	"The complete genome sequence of the lactic acid bacterium Lactococcus lactis ssp. lactis IL1403."
RL	Genome Ref. 11:731-753(2001).	RL	Genome Ref. 11:731-753(2001).
CC	-1- FUNCTION: REQUIRED FOR CELL SEPARATION DURING GROWTH (BY SIMILARITY).	CC	-1- FUNCTION: REQUIRED FOR CELL SEPARATION DURING GROWTH (BY SIMILARITY).
CC	-1- CATALYTIC ACTIVITY: Hydrolysis of the 1,4-beta-linkages between N-acetyl-D-glucosamine and N-acetylmuramic acid in peptidoglycan heteropolymers of the prokaryotic cell walls.	CC	-1- CATALYTIC ACTIVITY: Hydrolysis of the 1,4-beta-linkages between N-acetyl-D-glucosamine and N-acetylmuramic acid in peptidoglycan heteropolymers of the prokaryotic cell walls.
CC	-1- DOMAIN: THE LYSM REPEATS ARE THOUGHT TO BE INVOLVED IN PEPTIDOGLYCAN BINDING.	CC	-1- DOMAIN: THE LYSM REPEATS ARE THOUGHT TO BE INVOLVED IN PEPTIDOGLYCAN BINDING.
CC	-1- SIMILARITY: BELONGS TO FAMILY 73 OF GLYCOSYL HYDROLASES.	CC	-1- SIMILARITY: BELONGS TO FAMILY 73 OF GLYCOSYL HYDROLASES.

CC -1- SIMILARITY: CONTAINS 3 LYSM REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AE006264; AAK04370.1; -
 DR InterPro: IPR002901; Amidase-4.
 DR InterPro: IPR000423; Flag_FlgJ.
 DR InterPro: IPR002482; LysM.
 DR Pfam: PF01476; LysM; 3.
 DR Pfam: PF01832; Amidase 4; 1.
 DR PRINTS: PRO1002; FLAG_FlgJ.
 DR SMART: SM00047; LysM; 1.
 DR SMART: SM00257; LysM; 3.
 DR SMART: SM00257; LysM; 3.
 DR Hydrolase: Glycosidase; Bacteriolytic enzyme; Cell wall;
 DR Cell division; Septation; Signal; Repeat; Complete proteome.
 FT SIGNAL 1 57
 FT CHAIN 58 439 PROBABLE N-ACETYLTRANSFERASE.
 FT REPEAT 243 285 LYSM 1.
 FT REPEAT 323 365 LYSM 2.
 FT REPEAT 397 439 LYSM 3.
 FT DOMAIN 287 291 POLY-SER.
 FT DOMAIN 370 374 POLY-SER.
 SQ SEQUENCE 439 AA; 46592 MW; AE4551B4D04CB499 CRC64;
 Query Match 8.8%; Score 146.5; DB 1; Length 439;
 Best Local Similarity 23.6%; Pred. No. 0.0034;
 Matches 69; Conservative 39; Mismatches 107; Indels 77; Gaps 10;
 Oy 26 TTCILAGCASKPTYNST-----SGSGSHRTSGSGGLAIGCVITDSCGVNPR 72
 Db 184 TAAALGKATYAPNVCASLNRRIISQYNLTRFDGASAGTSNSG-STATNNNSNTSST 242
 Oy 73 YOVKQGVTVSKAORYGIMNREIGHNINLSYTYTGQWLTWLGDKVNERSTSSGVN 132
 Db 243 YTVSGDILWGLISQKTVSAOISANNDKST-VYIGQKLVLTSSSSNTSSTSSGN 301
 Oy 133 TAAHTSPVAVOSSRPVQOHPAVOKRTPPVVKKRTPPPVQOPAPVAPVT-----EA 188
 Db 302 SAGITTPPT-----SVTPAKPASQTTIKVKSQDILWGLSV 336
 Oy 189 PPAATSSGVMOFRYVGAATNPPVRRFGATVAGSTVTSNGMFFSGRGGDLINASAGTVI 248
 Db 337 KYKTTIAQLKSMNHLNSDTIPFGQMLIVSOSAGSSSSSTG-----SSSAST-- 382
 Oy 249 QADHMDGASIVYQHTNGFVSYSIHKDAOVYTGPTV-----RTGRIASMK 295
 Db 383 SSTSNSAAS-----NTSIH-----KVKGDILWGLSOKSGSPIASIK 420
 RESULT 10
 P60_LISMO STANDARD; PRT; 484 AA.
 ID P60_LISMO
 AC P21171; Q03493;
 DT 01-MAY-1991 (Rel. 18, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Protein p60 precursor (Invasion-associated protein).
 GN IAP OR LMO0582.
 OS *Listeria monocytogenes*.
 OC Bacteria; Firmicutes; Bacillales; *Listeriaceae*; *Listeria*.
 OX NCBI_TaxID=1639;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 28-49.
 RC STRAIN=EGD / Serovar 1/2a;
 RX MEDLINE=90256283; PubMed=2111287;
 RA Koehler S., Leimeister-Waechter M., Chakraborty T., Ioteppeich F.,
 RA Goebel W.;

RT "The gene coding for protein p60 of *Listeria monocytogenes* and its
 RT use as a specific probe for *Listeria monocytogenes*.";
 RL Infect. Immun. 58:1943-1950 (1990).
 RN [2]
 RP SEQUENCE FROM N.A., AND DISCUSSION OF SEQUENCE.
 RC STRAIN=Mackness / Serovar 1/2a;
 RX MEDLINE=93094153; PubMed=1459966;
 RA Bubert A., Kuhn M., Goebel W., Koehler S.;
 RT "Structural and functional properties of the p60 proteins from
 RT different *Listeria* species.";
 RL J. Bacteriol. 174:8166-8171 (1992).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=EGD-e / Serovar 1/2a;
 RX MEDLINE=21537279; PubMed=11679669;
 RA Glaser P., Frangoul L., Buchrieser C., Rusniok C., Anend A.,
 RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
 RA Charbit A., Chetouani F., Couve E., de Darivar A., Deloux P.,
 RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
 RA Eutian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
 RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
 RA Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kuypkat G.,
 RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjati H.,
 RA Nordstedt G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
 RA Remmel B., Rose M., Schluter T., Simoes N., Tierrez A.,
 RA Vazquez-Boland J.-A., Voss H., Wehlend J., Cossart P.,
 RT "Comparative genomics of *Listeria* species.";
 RL Science 294:849-852 (2001).
 CC -1- FUNCTION OF THIS MAJOR EXTRACELLULAR PROTEIN MAY BE INVOLVED IN THE
 CC INVASION OF NONPROFESSIONAL PHAGOCYtic CELLS BY *LISTERIA*.
 CC -1- DOMAIN: LYSM REPEATS ARE THOUGHT TO BE INVOLVED IN PEPTIDOGLYCAN
 CC BINDING.
 CC -1- SIMILARITY: BELONGS TO THE E. COLI NLP / *LISTERIA* P60 FAMILY.
 CC -1- SIMILARITY: CONTAINS 2 LYSM REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X52268; CAA36509.1; -
 DR EMBL: M80351; AAA25280.1; -
 DR EMBL: AL581975; CAC98661.1; ALT_INIT.
 DR PIR: A41487; A41487.
 DR L1actList; LMO00582; -
 DR MEROPS; C40.0PM; -
 DR InterPro: IPR002482; LysM.
 DR InterPro: IPR000064; SH3_P60.
 DR InterPro: IPR003646; SH3_Fac.
 DR Pfam; PF00877; NLP_P60; 2.
 DR Pfam; PF01476; LysM; 2.
 DR SMART; SM00257; LysM; 2.
 DR SMART; SM00287; SH3; 1.
 KW Repeat; Signal; Complete proteome.
 FT SIGNAL 1 27
 FT CHAIN 28 484
 FT DOMAIN 311 355 PROTEIN P60.
 FT REPEAT 30 72 19 X 2 AA TANDEM REPEATS OF T-N.
 FT REPEAT 203 245 LYSM 1.
 FT REPEAT 94 94 LYSM 2.
 FT VARIANT 167 167 S -> T (IN STRAIN MACKNESS).
 FT VARIANT 196 196 V -> I (IN STRAIN MACKNESS).
 FT VARIANT 326 331 MISSING (IN STRAIN MACKNESS).
 SQ SEQUENCE 484 AA; 50587 MW; 3CC0F90591E1E40F CRC64;
 Query Match 8.8%; Score 146.5; DB 1; Length 484;
 Best Local Similarity 20.6%; Pred. No. 0.0038;
 Matches 81; Conservative 46; Mismatches 121; Indels 145; Gaps 14;
 Oy 25 ITTCILAGCASKPTYNSTSGSGSHRTSGSGGLA--ISQVITD----- 65

Db 102 I I I S I K G - - T K Y T Y E T T E S N G M H K I T Y N D G K T G F V N K G L Y L D A V S T P A P T O E V K E T 159

Qy 66 - - S Q G V P - - - - - N R Q V N O G P T V S I A I O R Y G 89

Db 160 T T Q O A P A A E T K E V K O T T O A T T P A K A E I K E P V D Q N A T T H A V S G D I M A L S Y K G 219

Qy 90 L M R E I G H I N N L A N S T T I Y T G M L I M S G D L K V R R S I S G V N A T H P S P A A O S S R P V 149

Db 220 V S V O D L M S M N I S S - S I Y V G O K L A I - - - - - K O J A N T A I P A E A Y K T - - - - - 259

Qy 150 Q O H P A V O K E T P P V - - - - - V K K E P T P P V O Q P A V A P V T E A P A T O S S G M O 199

Db 260 - E A P A E K O A P V K E N T I N T N T A T T E K E T A O - - - - - O O T A K A P A T E A K A P A B A T I N T N A 315

Qy 200 F R A P V A - - - - - T N P V R R F G T A V A S I T V S N G M F S G R D G L I N A S N A G T V I O 249

Db 316 N K T N T N T N T N T N T N T N T P S K I N T I N S N T N T N S T N A N O S S N N N S S A S A I I A E 375

Qy 250 A D H N M D - - - - - G A S I Y I O H T N E F - V S S Y I H I D A O V K G D T 284

Db 376 A Q H I L G K A Y S W G N G E T P D C S G I T K Y V P A K A G I S L P R T S Q A O Y A S T T R I S E S A K A R G D L 435

Qy 285 V R T G O R I A S M K O P S G A L F E F - R I S R N G V V 315

Db 436 V - - - - - F P D Y G S G I S H V G I T V 451

```

RESULT 11
ID_ P60_LISTIV          STANDARD;          PRT;          524 AA.
AC_ 001837;
DT_ 01-APR-1993 (Rel. 25 Created)
DT_ 01-APR-1993 (Rel. 25, Last sequence update)
DT_ 16-OCT-2001 (Rel. 40, Last annotation update)
DE_ Protein p60 precursor (invasion-associated protein) .
GN_ IAP.
OS_ Listeria ivanovii.
OC_ Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX_ NCBI_TaxID=1638;
RN_ [1]
RP_ SEQUENCE FROM N.A.
RX_ MEDLINE=92384582; PubMed=1514809;
RA_ Rubert A., Koehler S., Goebel W.;
RT_ "the homologous and heterologous regions within the iap gene allow
RT_ genus- and species-specific identification of Listeria spp. by
RT_ polymerase chain reaction." ;
RN_ [2]
RN_ Appl. Environ. Microbiol. 58:2625-2632(1992).
RP_ DISCUSSION OF SEQUENCE.
RX_ MEDLINE=93094153; PubMed=1459966;
RA_ Rubert A., Kunh W., Goebel W., Koehler S.;
RT_ "Structural and functional properties of the p60 proteins from
RT_ different Listeria species." ;
RN_ J. Bacteriol. 174:8166-8171(1992).
CC_ -I- FUNCTION: THIS MAJOR EXTRACELLULAR PROTEIN MAY BE INVOLVED IN THE
CC_ INVASION OF NONPROFESSIONAL PHAGOCYtic CELLS BY LISTERIA.
CC_ -I- DOMAIN: LYSM REPEATS ARE THOUGHT TO BE INVOLVED IN PEPTIDOGLYCAN
CC_ BINDING.
CC_ -I- SIMILARITY: BELONGS TO THE E. COLI NLPC / LISTERIA P60 FAMILY.
CC_ -I- SIMILARITY: CONTAINS 3 LYSM REPEATS.
CC_ -----
CC_ This SWISS-PROT entry is copyright. It is produced through a collaboration
CC_ between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC_ the European Bioinformatics Institute. There are no restrictions on its
CC_ use by non profit institutions as long as its content is in no way
CC_ modified and this statement is not removed. Usage by and for commercial
CC_ entities requires a license agreement. See http://www.isb-sib.ch/announce/
CC_ or send an email to license@isb-sib.ch.
CC_ -----
DR_ EMBL; M80350; AAA25284.1; .
DR_ MEROPS; CA0.UFW; .
DR_ InterPro; IPR002482; LysM.

```

DR	InterPro:	IPR000064;	NPIC_P60.	
DR	InterPro:	IPR003646;	SH3_Bac.	
DR	Pfam:	PF00877;	NPIC_P60; 1.	
DR	Pfam:	PF01476;	LysM; 3.	
DR	SMART:	SM00257;	LysM; 3.	
DR	SMART:	SM00287;	SH3b; 1.	
KW	Repeat:	Signal.		
FT	SIGNAL	1		
FT	CHAIN	27		BY SIMILARITY.
FT	REPEAT	28	524	PROTEIN P60.
FT	REPEAT	30	72	LYSM 1.
FT	REPEAT	198	240	LYSM 2.
FT	REPEAT	314	356	LYSM 3.
SO	SEQUENCE	524 AA;	54159 MM;	E22D579429817144 CRC64;
Query Match				
Best local		8.7%;	Score 146;	DB 1;
Matches		78;	Conservative	40;
			Mismatches	95;
			Indels	104;
			Gaps	17;

Qy	25	ITTCILAGCASKPTYNSTSGSGHRTS	GSG	-----GLAIGSV-----	ITD	65
Db	100	IVTSLKGG	-TKTVVSTEANGWKISISGSKTGIVNGKVI	GIITVTSAPAEVKEETITQ		157
Qy	66	SQGP	-----	NRVQVQGDVTSKIAQRGLNM		92
Db	158	AQAPPAETKEPEVKOSTPATVL	PPAETKETETPAVDITASTVTS	SGDTIMALSKRGTSV		217
Qy	93	REIGHINNLSN	SYTIYTGOMLTLMWSGDKV	KERSISGCVNTAHTPSPVA	-VQSRPVPVQ	151
Db	218	QINISMNLS	SSS-SIYGOVLA	-----VQEA	-----AKTASPAAEKTETPAEK	262
Qy	152	H---	PAVQKPPPPVVVVKPPTPEPV	-VQCPAPAPVPTPEAFATGSSGVQFRYPVAT		207
Db	263	QVSPAVKENTQITTAKEVTP	POKOTQOAPQAAKP	-APAPAPVTNTNASSYTVKSGDT		321
Qy	208	NPVVRPFOTAVVASTVTS	-----NGMMPSGRGGDLINASN	-----AGTVIQADNM		254
Db	322	LSKI	-----ATTFGTVTSKIKALNG	-----LNSDNLQVGVLKVKGTVPANTNS		366
Qy	255	D-GASIVQHTNGFVSS	270			
Db	367	NSNATAPPTNTNNTSN	383			

RESULT 12						
PE0_LISSE						
ID_P60_LISSE	STANDARD:	PRT:	523	AA.		
AC	Q01838:					
DT	01-APR-1993	(Rel. 25, Created)				
DT	01-APR-1993	(Rel. 25, Last sequence update)				
DE	16-OCT-2001	(Rel. 40, Last annotation update)				
GN	Protein p60 precursor	(Invasion-associated protein).				
OS	Listeria seeligeri.					
OC	Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.					
OX	NCBI_TaxID=1640;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RX	MEDLINE=92384582;	PubMed=1514809;				
RA	Rubert A.; Koehler S.; Goebel W.;					
RT	"The homologous and heterologous regions within the iap gene allow					
RT	genus- and species-specific identification of Listeria spp. by					
RT	polymerase chain reaction.";					
RL	Appl. Environ. Microbiol. 58:2625-2632(1992).					
RN	[2]					
RP	DISCUSSION OF SEQUENCE.					
RX	MEDLINE=93094153;	PubMed=1459966;				
RA	Rubert A.; Kuhn W.; Goebel W.; Koehler S.;					
RT	"Structural and functional properties of the p60 proteins from					
RT	different Listeria species"	(1992)				
RL	J. Bacteriol. 174:8186-8171(1992)					
CC	-1- FUNCTION: THIS MAJOR EXTRACELLULAR PROTEIN MAY BE INVOLVED IN THE					
CC	INVASION OF NONPROFESSIONAL PHAGOCYTTIC CELLS BY LISTERIA					
CC	-1- DOMAIN: LISM REPEATS ARE THOUGHT TO BE INVOLVED IN PEPTIDOGLYCAN					

BINDING.
CC -1- SIMILARITY: BELONGS TO THE E.COLI NLPC / LISTERIA P60 FAMILY.
CC -1- SIMILARITY: CONTAINS 3 LYSM REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M80353; AAA25286.1; -
CC DR MEROPS; C40_UF; -
CC DR InterPro; IPR002482; LYSM.
CC DR InterPro; IPR000064; NLPC_P60.
CC DR InterPro; IPR003646; SH3_Fac.
CC DR Pfam; PF00877; NLPC_P60; 1.
CC DR Pfam; PF01476; LYSM; 3.
CC DR SMART; SM00257; LYSM; 3.
CC DR SMART; SM00287; SH3b; 1.
CC Repeat; Signal.
CC FT SIGNAL 1 27 BY SIMILARITY.
CC FT CHAIN 28 523 PROTEIN P60.
CC FT REPEAT 30 72 LYSM 1.
CC FT REPEAT 200 242 LYSM 2.
CC FT REPEAT 320 362 LYSM 3.
CC SQ SEQUENCE 523 AA; 53844 MW; 893412102D20DF6B4 CRC64;

Query Match 8.7%; Score 144.5; DB 1; Length 523;
Best Local Similarity 20.4%; Pred. No. 0.0057;
Matches 93; Conservative 58; Mismatches 131; Indels 173; Gaps 21;

QY 1 MNTVAINSQNGKPKRGL-----IFGVTTTCLGCAKPRYNTSGSGSR-TSG 52
DB 69 LQVTVASEKTEKYSATWLNVRSGAGVNSTVSLKQ3--TKVVESTEANGMKITYG 126
QY 53 SG-----GLAIGSOVITDSQGV----- 70
DB 127 ESKTGYVNGKYLGNVNTSPATPRVKEQETQAAPQOTKEVQAATPEKDAVET 186
QY 71 -----NRVYKQGDIVSKIAQRYGLNREIGHINNINSSYTYTGQMLTMSGD 119
DB 187 KTTAPAVDNTATTTHTVSGDITWALSVKYGASVDLMSWNLS--SIYVQ-----N 238
QY 120 LKVRERSISSGVNT-----AHTPSVAVOSSRPVQOHPAVKPPPVVVKKPPPT-- 171
DB 239 IAVKQ--SAAKNTAPKAEAKTEAPAEKQTAAPVVK--STNTSTTTVKKETTERKQ 292
QY 172 -----PPVQOAPAVAPV--TEAPFATGSSG----- 196
DB 293 TSTTAPAPQAAKAPAPAPVNTVNTASSTVYKSGDILGKIASFTGVSKIKALNGLTSDN 352
QY 197 -----VMQFRYPVAGTNPVRRFGTAIVAGSTVTSNGMWFSGRDGLINASNAGT----- 246
DB 353 LQVGVDLVKVKGIVPATNT-----NTAT-APTNTNN-----TSSSTSTSPSKNN 396
QY 247 -VIQADHMDGASITYQHT-----NGRVSSYTHIKAAQV--KTG 282
DB 397 NTNQSSSSSSASAIATIAEQKILGKAYSGWNGGPTTFDQSGP-TSYVFAQSGITLPRISG 455
QY 283 DTVRTGRIASMKNOPSGALFEF--RISRNGVYV 315
DB 456 AQYASTTKVSESAQPGDLVFPDYSGIAHVGIYV 490

RESULT 13

P60_LISGR
ID P60_LISGR STANDARD; PRT; 511 AA.
AC 001835;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Protein p60 precursor (Invasion-associated protein).
GN IAP.
OS Listeria grayi (Listeria murrayi).
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX NCBI_TaxId:1611;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=92384582; PubMed=1514809;
RA Hubert A., Koehler S., Goebel W.;
RT "The homologous and heterologous regions within the iap gene allow
RT genus- and species-specific identification of Listeria spp. by
RT polymerase chain reaction.";
RT Appl. Environ. Microbiol. 58:2625-2632(1992).
RN [2]
RP DISCUSSION OF SEQUENCE.
RX MEDLINE=93094153; PubMed=1459966;
RA Hubert A., Kuhn M., Goebel W., Koehler S.;
RT "Structural and functional properties of the p60 proteins from
RT different Listeria species.";
RL J. Bacteriol. 174:8166-8171(1992).
CC -1- FUNCTION: THIS MAJOR EXTRACELLULAR PROTEIN MAY BE INVOLVED IN THE
CC INVASION OF NONPROFESSIONAL PHAGOCYTTIC CELLS BY LISTERIA.
CC -1- DOMAIN: LYSM REPEATS ARE THOUGHT TO BE INVOLVED IN PEPTIDOGLYCAN
CC BINDING.
CC -1- SIMILARITY: BELONGS TO THE E.COLI NLPC / LISTERIA P60 FAMILY.
CC -1- SIMILARITY: CONTAINS 2 LYSM REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M80352; AAA25285.1; -
CC DR MEROPS; C40_UF; -
CC DR InterPro; IPR002482; LYSM.
CC DR InterPro; IPR000064; NLPC_P60.
CC DR InterPro; IPR003646; SH3_Fac.
CC DR Pfam; PF00877; NLPC_P60; 1.
CC DR Pfam; PF01476; LYSM; 3.
CC DR SMART; SM00257; LYSM; 3.
CC DR SMART; SM00287; SH3b; 1.
CC Repeat; Signal.
CC FT SIGNAL 1 27 BY SIMILARITY.
CC FT CHAIN 28 511 PROTEIN P60.
CC FT REPEAT 177 219 LYSM 1.
CC FT REPEAT 299 342 LYSM 2.
CC SQ SEQUENCE 511 AA; 53181 MW; 2FD0ED174E3810D0 CRC64;

Query Match 8.4%; Score 140.5; DB 1; Length 511;
Best Local Similarity 26.6%; Pred. No. 0.011;
Matches 47; Conservative 24; Mismatches 83; Indels 23; Gaps 5;

QY 73 YQVQGDIVSKIAQRYGLNREIGHINNINSSYTYTGQMLTMSGDLYKVERSSISGVN 132
DB 177 YKVSQDITWALSVKYGVPVQKLTBMNLS--SIYVQ-----TVAVEAAKAPPT 228
QY 133 TAHTPSVAVOSSRPVQOHPAVOKPPTPVV-VVKKPTPPPVVQOAPVAPVPTAPFA 191
DB 229 TVKQAPAKVAPKQEVQOTAPAKQKAPAKETVYKPAVSKPATPTAKPAVEQKAS 288
QY 192 TGSSGVQOFRYPVAGTNPVRRFGTAIVAGSTVTSNGMWFSGRDGLINASNAGTVI 248
DB 289 T-----PAVDTAATYKVN-----GDSLGIASLFFVSVADLTNNNNINATI 331

RESULT 14

P60_LISIN
ID P60_LISIN STANDARD; PRT; 467 AA.
AC 001836;
DT 01-APR-1993 (Rel. 25, Created)


```

CC -----
DR EMBL; M80354; AAA5281.1; ALT_SEQ.
DR EMBL; M80348; AAA5281.1; ALT_SEQ.
DR HSSP; P2391; 180G.
DR MEROPS; C40 UPB; -.
DR InterPro; IPR002482; LysM.
DR InterPro; IPR000064; NLP_C_P60.
DR Pfam; PF00877; NLP_C_P60_1.
DR Pfam; PF01476; LysM; 3.
DR SMART; SM00257; LysM; 3.
DR SMART; SM00287; SH3b; 1.
DR Repeat; Signal.
KM SIGNAL 1 27
FT CHAIN 28 524 BY SIMILARITY.
FT REPEAT 30 72 PROTEIN P60.
FT REPEAT 198 240 LYSM 1.
FT REPEAT 316 356 LYSM 2.
SQ SEQUENCE 524 AA; 54086 MW; 228AF5DC4A254C9 CRC64;

```

```

Query Match 8.2%; Score 137.5; DB 1; Length 524;
Best Local Similarity 22.9%; Pred. No. 0.017;
Matches 74; Conservative 37; Mismatches 103; Indels 109; Gaps 14;

```

```

QY 73 YQVKGQDTVSKIAQRYGLNWRIGHINNINSYTYITGQWLTMSGDKVERSSISGVN 132
DB 198 HTVKSQDITWALSVKYGASVDLMGMNLS-SIYVQ-----KIAVQS 242
QY 133 TAHTSPVAVQSSRPVQOHPVQKPT-PPVY-----VYKK-----PTPPPVQ 176
DB 243 AAKTAPAEVYKT-----EAPAVEKETSTPVKENTNTTVKKEVTTQTOTTKAPAQA 297
QY 177 QAPVAPVY-----TEAPATSSGVMQFRYPVGAATNPVRREGTAT----- 218
DB 298 KPAPAPAPVYNTASTYTVKSG-----DSIKANTGTISVKIKALNLTSDNL 349
QY 219 -----VAGSTVTSNGMFMFGSDGDLINASNAGT-----VIQADNMDC 257
DB 350 QVGTVLAKYGVPTNTNNSNTTAPTNTNNSNTSMTSPSKNTNTNNGSSNSASA 409
QY 258 SIVIOHT-----NGFVSYIHIDA--QVKTGPTVATGGRISM 294
DB 410 SALIAEAKHUKAYSWGNGPTTPDCSGF-TSYVFAKSGISLPTSGAQVASTRISBS 468
QY 295 KNPSGALFEF--RISNGVYV 315
DB 469 QAKGEDLVFPDGGSGIAHVGYV 491

```

RESULT 16

```

YIBP_HAEIN
ID YIBP_HAEIN STANDARD; PRT; 410 AA.
AC P4464;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein HI0756.
GN HI0756.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shiley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,

```

```

RA Gnehm C.L., McDonald L.A., Small K.V., Frazer C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd.";
RL Science 269:496-512(1995).
CC -I- SIMILARITY: STRONG, TO E.COLI YIBP.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by, and for commercial
CC entities requires a license agreement (See http://www.ibt-sib.ch/announce/
CC or send an email to license@ibt-sib.ch).
CC -----
CC EMBL; U32760; AAC22415.1; -.
DR MEROPS; M37.UPB; -.
DR TIGR; HI0756; -.
DR InterPro; IPR002886; Peptidase_M37.
DR Pfam; PF01551; Peptidase_M37; 1.
DR Hypothetical protein; Complete proteome.
SQ SEQUENCE 410 AA; 45983 MW; B28D93E813BCA9 CRC64;

```

```

Query Match 8.1%; Score 135; DB 1; Length 410;
Best Local Similarity 25.0%; Pred. No. 0.02;
Matches 63; Conservative 36; Mismatches 111; Indels 42; Gaps 10;

```

```

QY 74 QVKGQDTVSKIAQRYGLNWRIGHINNINSYTYITGQWLTMSGDKVRRSSISGVN 133
DB 188 QKKQQAALQKQOQH-----STLNELKNLALDQDQNT-----LKANQALRQETQR 236
QY 134 AHTSPVAVQSSRPV--QOHPVQKPTPPVVVKKPTTPPVVQOHPVAPVTEAPAT 192
DB 237 AEQAAREGKERERALLQKQAEKRT-----SKYQGTVERQGLN 278
QY 193 GSGGV---WQFRYPVGAATNPVRPFGTATVAGSTVTSNGMFGSDGDLINASNAGTVI 248
DB 279 STSGIGAAKQKQYSLPVSGS--ILHTFG--SIQAGEVWKGVAVIGASAGTPVKALIAAGRVI 334
QY 249 QADH-NMDGASIVIOHTNGFVSYIHIKDAVKTGPTVATGGRISMKN--QPSGALFE 305
DB 335 LAGTNGYGVNIVYHGETDLSLGFNOVASVKVQGLVASQVLAQVNTGTEISRSALY 393
QY 306 FRISNGVYVDP 317
DB 394 FGISRKGPVNP 405

```

RESULT 17

```

YEBP_HAEIN
ID YEBP_HAEIN STANDARD; PRT; 475 AA.
AC P4463;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein HI0409.
GN HI0409.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shiley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Frazer C.M., Smith H.O.,

```

RA Venter J.C.;
 RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
 RL Science 269:496-512 (1995).
 CC -1- FUNCTION: COULD BE INVOLVED IN CELL WALL DEGRADATION OR
 CC FORMATION (BY SIMILARITY).
 CC -1- SIMILARITY: STRONG, TO E. COLI Y8A.
 CC -1- SIMILARITY: TO STAPHYLOCOCCUS LYSOSTAPHIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U32724; AAC22068.1; -.
 DR MEROPS; M37; UPW; -.
 DR TIGR; H10409; -.
 DR InterPro; IPR002886; Peptidase_M37.
 DR Pfam; PF01551; Peptidase_M37; 1.
 KW Hypothetical protein; Cell wall; Complete proteome.
 SQ SEQUENCE 475 AA; 53255 MW; 143C10F92233939D CRC64;
 Query Match 7.8%; Score 132; DB 1; Length 475;
 Best Local Similarity 23.0%; Pred. No. 0.037;
 Matches 67; Conservative 37; Mismatches 83; Indels 104; Gaps 12;
 QY 72 RYGVKGD--TVSKIAQRYGLNMRERIGHNINL-----NSSYIT-----YTQG 111
 DB 217 RKEVLAGEIGNLSNLSREKGLDTRQISQSLNALQWQVSLKLLKKGQFALIVSRREYLD 276
 QY 112 WLLTLMGDLKYRERRSISGVNTAHTSPVAVQS-----SRPVQCHRA 154
 DB 277 KLT---GQGNVELRISSGGKXY-----AVQANGRYVNOGQETLKGARFPLQROAR 328
 QY 155 VQKPTPPVYVVKPTPPVYVQDPAPVAPVPTAEPATYSSGVQRRYPVGAATNPVYRRF 214
 DB 329 VSSPFPD-----NRHP----- 340
 QY 215 GATVAGSTVTSNGMFGSGDGLINASNAQVIYQADHNMDGAS--IVIOHTNGFVSSYI 272
 DB 341 ---VTGRIRPHKGVDPVSQGFPIAPADGTVKAVAYQAGAGRYVMLRHGREYQTVYM 396
 QY 273 HIRDAQVKTGDIYRTGQRILASMKNQ--PSGAAL--FEFRISRNGVYVDPPLV 320
 DB 397 HLSKSLVKAQQTIVKGERIALSGNTGISTGPHLHYEPHI--NGRAVAVPLTV 445
 RESULT 18
 ACMA LACLC STANDARD; PRT; 437 AA.
 AC Q48603; 052362;
 DT 16-OCT-2001 (Ref. 40, Created)
 DT 16-OCT-2001 (Ref. 40, Last sequence update)
 DT 16-OCT-2001 (Ref. 40, Last annotation update)
 DE Probable N-acetylglucosaminidase precursor (EC 3.2.1.17) (Peptidoglycan
 DE hydrolase) (Autolysin) (lysosyme).
 GN ACMA.
 OS Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
 OX NCBI_TaxId=1359;
 RN 11
 RP SEQUENCE FROM N.A.
 RC STRAIN=M61363;
 RX MEDLINE=95189733; PubMed=7883712;
 RA Buist G., Kok J., Leenhouts K.J., Dabrowska M., Venema G.,
 RA Haandrikman A.J.;
 RT "Molecular cloning and nucleotide sequence of the gene encoding the
 RT major peptidoglycan hydrolase of Lactococcus lactis, a muramidase
 RT needed for cell separation."
 RL J. Bacteriol. 177:1554-1563 (1995).

RN (12)
 RP SEQUENCE FROM N.A.
 RC STRAIN=2250;
 RX MEDLINE=21015606; PubMed=11131071;
 RA Govindasamy-Lucey S., Gopal P.K., Sullivan P.A., Pillidge C.J.;
 RT "Varying influence of the autolysin, N-acetyl muramidase, and the cell
 RT envelope protease on the rate of autolysis of six commercial
 RT Lactococcus lactis cheese starter bacteria grown in milk."
 RL J. Dairy Res. 67:585-596 (2000).
 CC -1- FUNCTION: HYDROLYSES THE CELL WALL OF L. LACTIS AND
 CC M. LYSOGENITICS. REQUIRED FOR CELL SEPARATION DURING GROWTH.
 CC CATALYTIC ACTIVITY: Hydrolysis of the 1,4-beta-linkages between N-
 CC heteropolymers of the prokaryotic cell walls.
 CC -1- DOMAIN: THE LYSM REPEATS ARE THOUGHT TO BE INVOLVED IN
 CC PEPTIDOGLYCAN BINDING.
 CC -1- SIMILARITY: BELONGS TO FAMILY 73 OF GLYCOSYL HYDROLASES.
 CC -1- SIMILARITY: CONTAINS 3 LYSM REPEATS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U17696; AAC3367.1; -.
 DR EMBL; AF036720; AAC93629.1; -.
 DR InterPro; IPR002901; Amidase_4.
 DR Pfam; PF01476; LysM; 3.
 DR Pfam; PF01832; Amidase_4; 1.
 DR SMART; SM00047; LysM; 3.
 DR SMART; SM00257; LysM; 3.
 KW Hydrolase, Glycosidase, Bacteriolytic enzyme, Cell wall;
 KW Cell division; Septation; Signal; Repeat.
 FT SIGNAL 1 57
 FT CHAIN 58 437
 FT REPEAT 245 287
 FT REPEAT 321 363
 FT REPEAT 395 437
 FT REPEAT 27 27
 FT REPEAT 31 31
 FT REPEAT 64 64
 FT REPEAT 155 155
 FT REPEAT 190 190
 FT REPEAT 231 231
 FT REPEAT 285 285
 FT REPEAT 316 316
 FT REPEAT 320 320
 FT REPEAT 368 368
 FT CONFLICT 368 368
 SQ SEQUENCE 437 AA; 46564 MW; 5C90563BD5DE28B CRC64;
 Query Match 7.8%; Score 130.5; DB 1; Length 437;
 Best Local Similarity 24.1%; Pred. No. 0.043;
 Matches 69; Conservative 39; Mismatches 119; Indels 59; Gaps 12;
 QY 26 TTCIILAGASRETVNST-----SGSGSHRTSGGGLAIGSOVITDQGVN- 71
 DB 184 TVALNGRYVYDSYASLANRIISQYNLRFQDASAGNTNSG---STTITNNNSGINS 240
 QY 72 ---RYGVKGDPTVSKIAQRYGLNMRERIGHNINNSYITYGQWLTMSGLKVERIS 128
 DB 241 SSTTYVSGDPLTWGISQRYGIVAQIOSANLAKST-IITGQKVLVGS---ASSTNS 295
 QY 129 SGVNTAHTSPVAVQSSRPVQDPAPVOKPTPPVYVVKPPT-PPVQDPAPVAPVPT 186
 DB 296 GGSNNS-----ASTPTTTSVTPA--KPTSQITVKKSGDPLVALSVKKTSTIAQKS 345
 QY 187 EAPFATGSGVQWQ---FRYPVGAATNPVVRFGTATVAGSTVTSNGMFGSGDGLINASN 243
 DB 346 WNLHLSDDITVIGNLIVSOSAAASNP-----STGSGSTATNNNSNSTSSNSNASHIKV 398

QY 244 AGTVIOADHNMDCASIVIOHTNGFVSYIHIKDAQVKTGDTVTRGQ 289
 DB 399 KGDITLWGLSOKSGSPI-----ASIKAMNHL-----SSDTLLIQ 432

RESULT 19

PRLB_ACHLY STANDARD; PRT; 374 AA.
 AC P27458;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Beta-lytic metalloendopeptidase precursor (EC 3.4.24.32) (Beta-lytic protease).
 OS Achromobacter lyticus.
 OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
 OC Achromobacter.
 OX NCBI_TaxID=224;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 196-220.
 RC STRAIN=M497-1;
 RX MEDLINE=91035265; PubMed=2228973;
 RA Li S.L., Norioka S., Sakiyama F.;
 RT "Molecular cloning and nucleotide sequence of the beta-lytic protease gene from Achromobacter lyticus.";
 RL J. Bacteriol. 172:6506-6511 (1990).
 CC -1- CATALYTIC ACTIVITY: Cleavage of N-acetyl[muramyl]-Ala, and of the
 CC inulin B chain at 23-Gly-|-Phe-24 > 18-Val-|-Cys(SO(3)H).
 CC -1- CONJUNCTION: BINDS ONE ZINC ION.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M23.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL, M60896; AAA21906.1; -.
 CC DR PIR, A37151; LYXYLY.
 CC DR MEROPS, M23.001; -.
 CC DR InterPro: IPR000841; BLTYCTASE.
 CC DR PRINTS: PR00933; BLTYCTASE.
 CC KM Hydrolase; Metalloprotease; Zinc; Zymogen; Signal.
 CC FT SIGNAL 1 24
 CC FT PROPEP 25 195
 CC FT CHAIN 196 374 BETA-LYTIC METALLOENDOPEPTIDASE.
 CC FT DISULFID 261 307 BY SIMILARITY.
 CC FT DISULFID 351 364 BY SIMILARITY.
 CC FT METNL 316 316 ZINC (POTENTIAL).
 CC FT METNL 318 318 ZINC (POTENTIAL).
 CC SQ SEQUENCE 374 AA; 40084 MW; 431E51B845DE14 CRC64;

Query Match 7.6%; Score 126.5; DB 1; Length 374;
 Best Local Similarity 25.9%; Pred. No. 0.068; Indels 49; Gaps 7;
 Matches 50; Conservative 29; Mismatches 65; Indels 49; Gaps 7;

QY 117 SGLDKVRRSSISGVNTAHTSPVAVOSRPVQOHAPVOKETPPVVVKKPTPPVVO 176
 DB 154 AGGRAARRRVPAGL-----RPVQHTAGGGGFGP-----LR 186
 QY 177 GPAPVAPVTEAPPTGSSGVQFPYVGAIVPVRRFGTAVAS-----TVTSNG 228
 DB 187 QCRP-----GRAVVSNGTLQFPFPGASWV---GGAHNTSGNGVPMSSLMRSRG 236
 QY 229 WPFSGRDLINASAGTVIOADHNMDCASIVIOHTNGFVSYIHIKDAQVKTGDTVTRGQ 288
 DB 237 GKGSGNQNGVWVSASASGFKR--HSSCPAEIV--HTGWSITYYHLNMIQNTGANVSMN 292
 QY 289 ORIASMKQPSGA 301

DB 293 TAINPANTQAOA 305

RESULT 20

GUNA_CALSA STANDARD; PRT; 1742 AA.
 AC P22534;
 DT 01-AUG-1991 (Rel. 19, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE Endoglucanase A precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase A) (Cellulase A).
 GN CEBA.
 OS Caldocellum saccharolyticum (Caldicellulosiruptor saccharolyticus).
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;
 OC Caldicellulosiruptor.
 OX NCBI_TaxID=44001;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=95336703; PubMed=7612247;
 RX Te'O V.S., Saul D.J., Bergquist P.L.;
 RA "Cela, another gene coding for a multidomain cellulase from the
 RT extreme thermophile Caldocellum saccharolyticum.";
 RL Appl. Microbiol. Biotechnol. 43:291-296 (1995).
 RN [2]
 RP SEQUENCE OF 1516-1742 FROM N.A.
 RC MEDLINE=91247819; PubMed=2039230;
 RX Leuchli E., Usamat N.B., Grayling R.A., Love D.R., Bergquist P.L.;
 RA "Cloning, sequence analysis, and expression in Escherichia coli of a
 RT gene coding for a beta-mannanase from the extremely thermophilic
 RL bacterium 'Caldocellum saccharolyticum'.";
 RL Appl. Environ. Microbiol. 57:694-700 (1991).

 CC -1- FUNCTION: THE N-TERMINAL DOMAIN OF CELA ENCODES FOR AN
 CC ENDOGLUCANASE ACTIVITY ON CARBOXYMETHYLCELLULOSE. THE C-TERMINAL
 CC DOMAIN PROBABLY ACT SYNERGISTICALLY TO HYDROLYZE CRYSTALLINE
 CC CELLULOSE.
 CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
 CC linkages in cellulose.
 CC -1- PTM: THE LINKER REGION (ALSO TERMED "HINGE") MAY BE A POTENTIAL
 CC SITE FOR PROTEOLYSIS.
 CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY
 CC E (FAMILY 9 OF GLYCOSYL HYDROLASES).
 CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY
 CC L (FAMILY 48 OF GLYCOSYL HYDROLASES).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL, L32742; AAA91086.1; -.
 CC DR EMBL, M36063; AAA72860.1; -.
 CC DR PIR, A43745; A43745.
 CC DR HSSP, P26221; ITP4.
 CC DR InterPro: IPR001956; CBD_3.
 CC DR InterPro: IPR001701; GH_9.
 CC DR InterPro: IPR000556; Glyco_hydro_48.
 CC DR Pfam: PF00759; Glyco_hydro_9; 1.
 CC DR Pfam: PF00942; CBM_3; 3.
 CC DR Pfam: PF02011; Glyco_hydro_48; 1.
 CC DR PRINTS: PR00844; GLHYDRLASE48.
 CC DR ProDom: PD011903; Glyco_hydro_48; 1.
 CC DR ProDom: PD011903; Glyco_hydro_48; 1.
 CC DR ProSite: PS00592; GLYCOSYL_HYDROL_F9_1; 1.
 CC DR ProSite: PS00698; GLYCOSYL_HYDROL_F9_2; 1.
 CC KM Cellulose degradation; Hydrolase; Glycosidase; Signal; Repeat.
 CC FT SIGNAL 1 23
 CC FT CHAIN 24 1742 ENDOGLUCANASE A.

```

FT DOMAIN 24 642 CATALYTIC 1.
FT DOMAIN 643 700 LINKER ("HINGE") (PRO-THR BOX).
FT DOMAIN 701 857 CELLULOSE-BINDING (BY SIMILARITY).
FT DOMAIN 858 903 LINKER ("HINGE") (PRO-THR BOX).
FT DOMAIN 904 1060 CELLULOSE-BINDING (BY SIMILARITY).
FT DOMAIN 1061 1112 LINKER ("HINGE") (PRO-THR BOX).
FT DOMAIN 1113 1142 CATALYTIC 2.
FT ACT SITE 396 396 BY SIMILARITY.
FT ACT SITE 434 434 BY SIMILARITY.
FT ACT SITE 443 443 BY SIMILARITY.
FT CONFLICT 1545 1545 T->A (IN REF. 2).
SQ SEQUENCE 1742 AA; 193696 MW; 3F069A2123EED07 CRC64;

Query Match
Best Local Similarity 23.5%; Pred. No. 0.56;
Matches 76; Conservative 51; Mismatches 130; Indels 66; Gaps 17;

7 INSGNQKPIKRLGLI-----FGVITTCILAGCASKPTVNSTSGSGSHRTSG-----SGGL 56
502 VNNQSGWPARATNKLKFRFVDSLILKAGVSPNQLTSTVYNQAGKVGYPVWDSHNI 561
57 AIGSGVITDSQGV---PNRYQVKGDTVSKIAQRYGLNMRREIGHINNIN-----S 103
562 ---YIIVDFGTGLIYPGGQDKYKKEVQFRIAPQNVQW-----DNSNDYSFODIKGV 612
104 SYTIYTGOMLTMSGDLVY-REKSISSGVN-----TAHPSPVAVQSSPPVQGHAVQKP 158
613 SSSVKTATKIPYXDDIKWGEPEGTSVSPPTASVPTPTPTPTPTPTPTPTPTPTPTPT 671
159 TPVVVVVKKPTP---TPPVVQCPAPVAPVTEAPFAPGSSGVMGFRYPVATNVPVRF 214
672 IFTVATPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 726
215 -----GTAATVAGSVTNSGMPFSGRGGDLINASNAQVITYQADHANMDGASITYQHTNGFV 268
727 LKVNSSGSSSIDLSRVTIR-YWYT-----VDGERQSAI-SDMAQIGASNV---TFKTV 775
269 --SSYIHKDAQVKTGDTVTRTQ 289
776 KLSSSVSGADYVLEIGFKSGAQ 798

RESULT 21
MUR2_ENTHR STANDARD; PRT; 666 AA.
ID MUR2_ENTHR
AC P39046;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Muramidase-2 precursor (EC 3.2.1.17) (1,4-beta-N-acetylmuramoylhydrolase) (Peptidoglycan hydrolase) (Pg-hydrolase-2) (Lysozyme).
OS Enterococcus hirae.
OC Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.
OX NCBI_TaxID:1354;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 50-73.
RC STRAIN=ATCC 9790;
RC MEDLINE=92165737; PubMed=1347040;
RA Chu C.-P., Kariyama R., Daneo-Moore L., Shockman G.D.;
RT "Cloning and sequence analysis of the muramidase-2 gene from Enterococcus hirae."
RT J. Bacteriol. 174:1619-1625 (1992).
RN [2]
RP FUNCTION.
RC STRAIN=ATCC 9790;
RC MEDLINE=89327152; PubMed=2753858;
RA Dolinger D.L., Daneo-Moore L., Shockman G.D.;
RT "The second peptidoglycan hydrolase of Streptococcus faecium ATCC 9790 covalently binds penicillin."
RT J. Bacteriol. 171:4355-4361 (1989).
RN [3]
RP FUNCTION.

```

```

RC STRAIN=ATCC 9790;
RA Del Mar Lleo M., Canepari P., Satta G.;
RT "Thermosensitive cell growth mutants of Enterococcus hirae that elongate at non-permissive temperature are stimulated to divide by parental autolytic enzymes."
RT J Gen. Microbiol. 139:3099-3117 (1993).
CC -1- FUNCTION: MAY WORK IN CONCERT WITH AND POTENTIATE THE PROCESSIVE HYDROLYTIC ACTION OF MURAMIDASE-1, WHICH REQUIRES BINDING OF THE ENZYME TO NONREDUCING ENDS OF GLYCAN CHAINS. HYDROLYSIS IN THE MIDST OF GLYCAN CHAINS WOULD INCREASE THE NUMBER OF BINDING SITES FOR MURAMIDASE-1. MAY FUNCTION IN FACILITATING SEPTUM FORMATION AND CELL SEPARATION. ACTIVE ON M. LUTERIS CELL WALLS AND ON E. HIRAE CELL WALL FRACTIONS, BUT NOT ACTIVE ON E. HIRAE INTACT CELL WALLS. CAN COVALENTLY BIND PENICILLIN of the 1,4-beta-linkages between N-CATALYTIC ACTIVITY; Hydrolysis of the 1,4-beta-linkages between N-acetyl-D-glucosamine and N-acetylmuramic acid in peptidoglycan. CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- DOMAIN: LYSM REPEATS ARE THOUGHT TO BE INVOLVED IN PEPTIDOGLYCAN BINDING.
CC -1- SIMILARITY: BELONGS TO FAMILY 73 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M7639; AAA24776.1;
DR PIR: A42296;
DR HSSP: P26229; 15MD.
DR InterPro: IPR002901; Amidase_4.
DR InterPro: IPR002482; LysM.
DR Pfam: PF01476; LysM; 6.
DR SMART: SM00832; Amidase_4; 1.
DR SMART: SM00047; LY22; 1.
DR SMART: SM00257; LysM; 6.
DR Hydrolase; Glycosidase; Bacteriolytic enzyme; Cell wall;
KM Cell division; Septation; Repeat; Signal.
FT SIGNAL 1 49
FT CHAIN 50 666 MURAMIDASE-2.
FT REPEAT 257 299 LYSM 1.
FT REPEAT 338 380 LYSM 2.
FT REPEAT 414 456 LYSM 3.
FT REPEAT 489 531 LYSM 4.
FT REPEAT 565 607 LYSM 5.
FT REPEAT 623 665 LYSM 6.
SQ SEQUENCE 666 AA; 70670 MW; FFOAYFAFCD810BA3 CRC64;

Query Match
Best Local Similarity 31.6%; Pred. No. 0.21;
Matches 36; Conservative 21; Mismatches 40; Indels 17; Gaps 4;

31 AGCAKSPPTVNSTSG---SGSHRTSGSGGLAIGSVITDSQGVPNRYQVKGDTVSKIAQR 87
DB AGSSTTNGNNAASGNTGNTSGTGOATGA-----KYTKSGDSQWKIND 352
OY 68 YELNREIGHINNINSYTIYTGOMLTMSGDLKVRERSISSGVTAHTPSPVA 141
DB 353 HOISNQLLENNINKNF-VYFGQQLVVKSSSSA--SGSTNSTGTNTSSNTA 403

RESULT 22
LAI7_YEAST STANDARD; PRT; 633 AA.
ID LAI7_YEAST
AC Q12446;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Proline-rich protein Lai7.

```

GN LAS17 OR YOR181W.
 OS Saccharomyces cerevisiae (Baker's Yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 NCBI_TaxID=4932;
 RN (1)
 RP SEQUENCE FROM N.A.
 RA Toh-E.A.;
 RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
 RN (2)
 RP SEQUENCE FROM N.A.
 RA Hughes B., Pohl T.M.;
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: TO S.POMBE WSP1.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC
 DR EMBL; D78487; BAA11386.1; -
 DR EMBL; Z75089; CAA9390.1; -
 DR SGD; S0005707; LAS17.
 DR InterPro; IPR000697; RanBpl_WASP.
 DR InterPro; IPR001960; WH1.
 DR InterPro; IPR003124; WH2.
 DR Pfam; PF00568; WH1; 1.
 DR Pfam; PF02205; WH2; 1.
 DR SMART; SM00461; WH1; 1.
 DR SMART; SM00246; WH2; 1.
 DR DOMAIN 185 190 POLY-PRO.
 FT DOMAIN 323 329 POLY-PRO.
 FT DOMAIN 342 348 POLY-PRO.
 FT DOMAIN 352 358 POLY-PRO.
 FT DOMAIN 385 391 POLY-PRO.
 FT DOMAIN 427 431 POLY-PRO.
 FT DOMAIN 470 474 POLY-PRO.
 FT DOMAIN 503 507 POLY-PRO.
 FT DOMAIN 520 526 POLY-PRO.
 SQ SEQUENCE 633 AA; 67571 MW; 448835563AA2645 CRC64;
 Query Match 7 1%; Score 118; DB 1; Length 633;
 Best Local Similarity 23.0%; Pred. No. 0.47;
 Matches 65; Conservative 34; Mismatches 104; Indels 80; Gaps 12;
 QY 37 PTYNSTSGSG-----SHRTSGSGGLAIGSQVITDSOG----- 68
 DB 329 PMRTTSSGVRRLPAPPPRRGRAPPPRRHRTVNTLNSGNSLIPQATGRGRAPAP 388
 QY 69 -VPNRQYKQDQTVSKIAQRYGLNREIGHINLNS-----YTYTGQMLTMSG 118
 DB 389 PPRASRPTPVNTMQONPQYVNSNRPFQYTNSSNPPPPVTFNTLFPQ-WTAATG 447
 QY 119 DLKV-----KRSISSGVNTAHTSPYAVGSSRPVQOHAVQCPTEPVVVKETPTP 172
 DB 448 QPAVLPONTQAPSOATVPPVAPPPPPASLQSQSIQAPASPIPIPTLPSTTSAAPPPP 507
 QY 173 P-VVQOP---APVAPVTEAPFATGSSGVNQFRYPVATNVFVRRFGATVAGSTVSN 227
 DB 508 AFLTQPPSGAPAPPPPPQMP-ATSTG-----GGSFARTT 543
 QY 228 GMMFSGRDGDLIN---ASNAGTVIQADHN-MDGASIVTQHTNG 266
 DB 544 G--DAGRDALLASIRGAGGIGALRKVDKSQLDKPVLQLQEARG 584

AC Q9Y5U6;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Reticulon 4 (Neurite outgrowth inhibitor) (Nogo protein) (Foccen)
 DE (Neuroendocrine-specific protein) (NSP) (Neuroendocrine specific
 DE protein C homolog) (RTN-x) (Reticulon 5) (My043 protein).
 GN RTN4 OR NOGO OR ASY OR KIAA0886.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NCBI_TaxID=9606;
 RN (1)
 RP SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).
 RX MEDLINE=20129242; PubMed=10667780;
 RA Prinija R., Moore S.E., Vinson M., Blake S., Morrow R., Christie G.,
 RA Michalovich D., Simmons D.L., Walsh F.S.;
 RL "Inhibitor of neurite outgrowth in humans.";
 RL Nature 403:383-384(2000).
 RN (2)
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
 RC TISSUE=Brain;
 RX MEDLINE=21010696; PubMed=11126360;
 RA Tagami S., Eguchi Y., Kinoshita M., Takeda M., Tsujimoto Y.;
 RL "A novel protein, RTN-XS, interacts with both Bcl-XL and Bcl-2 on
 RL endoplasmic reticulum and reduces their anti-apoptotic activity.";
 RL Oncogene 19:5736-5746(2000).
 RN (3)
 RP SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).
 RX MEDLINE=20237542; PubMed=10773680;
 RA Yang J., Yu L., Bi A.D., Zhao S.-Y.;
 RL "Assignment of the human reticulon 4 gene (RTN4) to chromosome
 RL 2p14-->2p13 by radiation hybrid mapping";
 RL Cytogenet. Cell Genet. 88:101-102(2000).
 RN (4)
 RP SEQUENCE FROM N.A. (ISOFORM 4).
 RA Jin W.-L., Ju G.;
 RL "Developmentally-regulated alternative splicing in a novel Nogo-A";
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 RN (5)
 RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 3).
 RC TISSUE=Placenta, and Skeletal muscle;
 RA Ito T., Schwartz S.M.;
 RL "Cloning of a member of the reticulon gene family in human.";
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 RN (6)
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RC TISSUE=Fibroblast;
 RA Yutendo M.;
 RL "Isolation of a cell death-inducing gene";
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
 RN (7)
 RP SEQUENCE FROM N.A. (ISOFORM 3).
 RC TISSUE=Plutitary;
 RA Song B., Peng Y., Zhou J., Huang Q., Dai M., Mao Y.M., Yu Y., Xu X.,
 RA Luo B., Hu R., Chen J.;
 RL "Human neuroendocrine-specific protein C (NSP) homolog gene";
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 RN (8)
 RP SEQUENCE FROM N.A. (ISOFORM 3).
 RA Gu J.R., Wan D.F., Zhao X.T., Zhou X.M., Jiang H.Q., Zhang P.P.,
 RA Qin W.X., Huang Y., Qiu X.K., Qian L.F., He L.P., Li H.N., Yu Y.,
 RA Yu J., Han L.H.;
 RL "Novel human cDNA clone with function of inhibiting cancer cell
 RL growth.";
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
 RN (9)
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC TISSUE=Brain;
 RX MEDLINE=99156230; PubMed=10048485;
 RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hirosewa M.,
 RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
 RL "Prediction of the coding sequences of unidentified human genes. XII.

RA Kageyama R.;
 RT "Structure, chromosomal locus, and promoter analysis of the gene
 RT autoregulation through the multiple N box elements.";
 RT J. Biol. Chem. 269:5150-5156(1994).
 CC -1- FUNCTION: TRANSCRIPTIONAL REPRESSOR OF GENES THAT REQUIRE A BHLH
 CC PROTEIN FOR THEIR TRANSCRIPTION. MAY ACT AS A NEGATIVE REGULATOR
 CC OF MYOGENESIS BY INHIBITING THE FUNCTIONS OF MYOD AND ASH1.
 CC -1- SUBUNIT: TRANSCRIPTION REPRESSION REQUIRES FORMATION OF A COMPLEX
 CC WITH A CO-REPRESSOR PROTEIN (GROUCHO).
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- TISSUE SPECIFICITY: EXPRESSED AT HIGH LEVELS IN UNDIFFERENTIATED
 CC NEURAL PRECURSOR CELLS, BUT THE LEVEL OF EXPRESSION DECREASES AS
 CC NEURAL DIFFERENTIATION PROCEEDS.
 CC -1- DOMAIN: HAS A PARTICULAR TYPE OF BASIC DOMAIN (PRESENCE OF A
 CC HELIX-INTERDISPTING PROLINE) THAT BINDS TO THE N-BOX (CAGCAG),
 CC RATHER THAN THE CANONICAL E-BOX (CANNTG).
 CC -1- DOMAIN: THE CARBOXYL-TERMINAL WRW MOTIF IS A TRANSCRIPTIONAL
 CC REPRESSION DOMAIN NECESSARY FOR THE INTERACTION WITH GROUCHO, A
 CC TRANSCRIPTIONAL CO-REPRESSOR RECRUITED TO SPECIFIC TARGET DNA BY
 CC HAIRY-RELATED PROTEINS.
 CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
 CC TRANSCRIPTION FACTORS. "HAIRY" SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL, D16464; BAA03931.1; -
 DR PIR, A53336; A53336.
 DR TRANSFAC; T01649; -
 DR MGD; MGI:104853; Hes1.
 DR InterPro; IPR001092; HLH basic.
 DR InterPro; IPR003650; Orange.
 DR Pfam; PF00010; HLH; 1.
 DR SMART; SM00353; HLH; 1.
 DR SMART; SM00511; ORANGE; 1.
 DR PROSITE; PS00038; HLH_1; 1.
 DR PROSITE; PS00888; HLH_2; 1.
 KM Transcription regulation; DNA-binding; Nuclear protein; Repressor.
 FT DNA BIND 35 47
 FT DOMAIN 48 92 BASIC DOMAIN
 FT DOMAIN 156 248 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
 FT DOMAIN 251 273 PRO-RICH.
 FT DOMAIN 277 280 SER/THR-RICH.
 FT DOMAIN WRW MOTIF (REQUIRED FOR ACTIVITY)
 FT DOMAIN (BY SIMILARITY).
 SQ SEQUENCE 282 AA; 29749 MW; 86C7700C5EF7DA26 CRC64;
 Query Match 7.0%; Score 116.5; DB 1; Length 282;
 Best Local Similarity 27.4%; Pred. No. 0.24;
 Matches 43; Conservative 15; Mismatches 42; Indels 57; Gaps 8;
 QY 126 SISSGVNT-----AHTSPVA-VQSSRPVQGHAVOKRPVYVVKKPP----- 169
 DB 126 STCEGVNTEVTRLLIGHLANCMTQINAMTYPGQAHPALQAPPP-----PPSGAPGQHA 180
 QY 170 ----PTPPVQOPAVAPVPTTEAPFATGS-----SGVMQFRYPVGA- 206
 DB 181 PFAPPPPLVPIPGAAPPSPGAPCKLSQAGEAKVFGQFVAPDQGFALPNAF 240
 QY 207 --TNTVVRFF-----GTAT-----VAGSTVTSNGMW 230
 DB 241 AHSGVPIPVYTSNSGTSVGPNAVSPSSGSLTSDMW 277
 RESULT 25
 NLPD_BUCAL STANDARD; PRT; 334 AA.
 AC P57493;

DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Protein nlpd/ltpb homolog.
 GN BU418.
 OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
 OS symbiotic bacterium).
 OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
 OX NCBI_TaxId=118099;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Tokyo 1998;
 RX MEDLINE=20445173; PubMed=10993077;
 RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
 RT "Genome sequence of the endocellular bacterial symbiont of aphids
 RT Buchnera sp. APS.";
 RL Nature 407:81-86(2000).
 CC -1- SIMILARITY: BELONGS TO THE E. COLI NLPD / HAEMOPHILUS LPPB FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL, AP001119; BABJ3116.1; -
 DR InterPro; IPR002482; LysM.
 DR InterPro; IPR002886; Peptidase_M37.
 DR Pfam; PF01476; LysM; 1.
 DR Pfam; PF01551; Peptidase_M37; 1.
 DR SMART; SM00257; LysM; 1.
 KM Complete proteome.
 SQ SEQUENCE 334 AA; 39788 MW; 4E4BD2117FD6977 CRC64;
 Query Match 6.9%; Score 116; DB 1; Length 334;
 Best Local Similarity 21.1%; Pred. No. 0.32;
 Matches 54; Conservative 45; Mismatches 131; Indels 26; Gaps 7;
 QY 73 YVKGQDTVSKIAQRYGLNREIGHINNLSYTYTGQMTLWSGDIKVERSSISGVN 132
 DB 91 YIVKSKDTMYIAKNSGVNYHELKFNKIKPKYIIIGQ--KIWGDPLDKNNDCSI- 147
 QY 133 TAHPSPPVAOSSSRPVQGHAVOKRPVYVVKKPPPTPPVQOPAVAPV----PVTEA 188
 DB 148 -----INLEKNSIKQHNSC-----VVFKNLNIKEFLKDNITKTKICEFCIKI 192
 QY 189 PFATGSGVMQFRYPVGTATNPVRRFGATVAGSTVTSNGMFGSDGLINASVAGTVI 248
 DB 193 KKNNSLKLKRFNSNMKSWVKK-KTKYIYIDKLAKRIEITFGQAPVFSTAGGVV 251.
 QY 249 QADH--NMDGASIVIOHTNGFVSSYIHKDAQVKTGDTVRTGRIASN-KNPSGALFE 305
 DB 252 FVTMLFKKYGILLITIKHQYLSIYAFNNSVLYKEDRVYKNOQIATWGLSSDTIARLY 311
 QY 306 FRISANGYVDPVTVL 321
 DB 312 FEIRYLGESINPLSL 327
 RESULT 26
 PREL_LYCES
 ID PREL_LYCES STANDARD; PRT; 346 AA.
 AC Q00451;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE 36.4 kDa proline-rich protein.
 GN TRP-R1.
 OS Lycopersicon esculentum (Tomato).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

```

OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
RN NCBI_TaxID=4081;
RP SEQUENCE FROM N.A.
RC STRAIN=cv VPNT Cherry; Tissue=fruit;
RX MEDLINE=92119262; PubMed=1731999;
RA Sales Y., Machin R., Kengsbuch D., Grunseem W., Barg R.;
RT DNA sequence of the tomato fruit expressed proline-rich protein gene
RT Trp-F1 reveals an intron within the 3 untranslated transcript.
RN Plant Mol. Biol. 18:407-409(1992).
RL (2)
RC SEQUENCE OF 34-346 FROM N.A.
RP STRAIN=cv. Arava; PubMed=1868217;
RX MEDLINE=91329722; PubMed=1868217;
RA Sales Y., Machin R., Grunseem W., Barg R.;
RT "Sequence coding for a novel proline-rich protein preferentially
RT expressed in young tomato fruit."
RL Plant Mol. Biol. 17:149-150(1991).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X61395; CAA43666.1; -
DR EMBL: X57076; CAA40361.1; -
DR HSSP: P24337; 1HYP
DR InterPro: IPR001768; Try/amy1_inhbr.
DR Pfam: PF00224; TRYF_1; 36375 MW; 60458452FEF16E0 CRC64;
SQ SEQUENCE 346 AA; 36375 MW; 60458452FEF16E0 CRC64;

Query Match 6.9%; Score 116; DB 1; Length 346;
Best Local Similarity 39.8%; Pred. No. 0.33; Indels 12; Gaps 5;
Matches 33; Conservative 6; Mismatches 12;

QY 136 TPSPVAVQSSRPVQGHAPAVQKPTPPV--VYKKP-TTPPVVQGPAPVAPVTPAPPA 191
DB 158 TTPPPPVV--HPPVTPKP--PSPTPVSPPIVPTTPPVSPPIITPPIVSPPV 212
QY 192 TGSSGVQGFY---PVGATNPV 211
DB 213 PNPVVVIPPVPSPPVTPPIV 235

RESULT 27
ID_NLDP YEREN STANDARD; PRT; 97 AA.
AC P47764;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Lipoprotein nlpd (fragment).
GN NLDP
OS Yersinia enterocolitica.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Yersinia
OC NCBI_TaxID=630;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=W1024 / Serotype O:9;
RX MEDLINE=95247270; PubMed=7729893;
RA Irlate M., Stahler I., Cornells G.R.;
RT "The rps gene from Yersinia enterocolitica and its influence on
RT expression of virulence factors."
RL Infect. Immun. 63:1840-1847(1995).
CC -1- SUBCELLULAR LOCATION: Attached to the inner membrane by a lipid
CC anchor (Potential).
CC -1- SIMILARITY: BELONGS TO THE E. COLI NLDP / HAEMOPHILUS LPPB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U16152; AAC43390.1; -
DR MEROPS: M37 UPW; -
DR InterPro: IPR002886; Peptidase M37.
DR InterPro: IPR000437; Pock_lipodrot.
DR Pfam: PF01551; Peptidase M37; 1
DR PROSITE: PS00013; PROKAL_LIPOPROTEIN, PARTIAL.
KW Inner membrane; Lipoprotein.
FT NON_TER
SQ SEQUENCE 97 AA; 10667 MW; 9E194F9A13D0838E CRC64;

Query Match 6.9%; Score 115; DB 1; Length 97;
Best Local Similarity 30.4%; Pred. No. 0.093; Indels 2; Gaps 1;
Matches 28; Conservative 21; Mismatches 41;

QY 232 SGKDDLLINASNAQTIVQADHNMDS--IVIQHTNGFVSSYIHKAQVKTGTVRTGQ 289
DB 3 AGSRQQPFATANGVVVAGNALGGYGNLIIKHNDVLSAYVHNDTWLVREQEVKAGQ 62
QY 290 RIASKKNQPSGALFEFRISRNQVVDPLTVL 321
DB 63 KIATWGSTGTSVRLHFRIYKGVNPLRYL 94

RESULT 28
ID_MOZ_HUMAN STANDARD; PRT; 2004 AA.
AC Q92794;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Monocytic leukemia zinc finger protein (Zinc finger protein 220).
GN ZNF220 OR MOZ.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=96376968; PubMed=8782817;
RA Borrow J., Stanton V.P. Jr., Andresen J.M., Becher R., Behm F.G.,
RA Chaganti R.S.K., Civan C.I., Distche C., Dube I., Fritsch A.M.,
RA Horman D., Mittleman F., Volinia S., Watmore A.E., Housman D.E.;
RT "The translocation t(8;16)(p11;p13) of acute myeloid leukaemia fuses
RT a putative acetyltransferase to the CREB-binding protein."
RL Nat. Genet. 14:33-41(1996).
CC -1- FUNCTION: MAY REPRESENT A CHROMATIN-ASSOCIATED ACETYLTRANSFERASE.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- DISEASE: PARTICIPATES IN A T(8;16)(P11;P13) CHROMOSOMAL
CC TRANSLOCATION THAT PRODUCES A MOZ-CBP CHIMERA OBSERVED IN THE
CC M4/M5 SUBTYPE OF ACUTE MYELOID LEUKEMIA.
CC -1- SIMILARITY: CONTAINS 2 PHD-TYPE ZINC FINGERS.
CC -1- SIMILARITY: BELONGS TO THE MOZ (SAS/MOZ) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U47742; AAC50662.1; -
DR Genew: HGNC:13013; ZNF220.
DR MIM: 601408; -
DR InterPro: IPR001386; Histone_H1/H5.
DR InterPro: IPR002717; MOZ_SAS.

```

```

DR InterPro; IPR001965; Znf_PHD.
DR Pfam; PF00628; PHD; 2.
DR Pfam; PF01853; MOZ_SAS; 1.
DR SMART; SMO0526; H15; 1.
DR SMART; SMO0249; PHD; 2.
DR PROSITE; PS01359; ZF_PHD_1; 1.
DR PROSITE; PS50016; ZF_PHD_2; 2.
KW Proto-oncogene; Chromosomal translocation; Zinc-finger; Repeat;
KW Nuclear protein.
FT ZN_PING 206 265 PHD-TYPE 1.
FT ZN_PING 259 313 PHD-TYPE 2.
FT DOMAIN 371 379 POLY-SER.
FT ZN_PING 538 560 C2HC-TYPE.
FT DOMAIN 788 801 POLY-GLU.
FT DOMAIN 989 995 POLY-GLU.
FT DOMAIN 1019 1026 POLY-ARG.
FT DOMAIN 1069 1078 POLY-GLU.
FT DOMAIN 1147 1150 POLY-LYS.
FT DOMAIN 1221 1242 GLU-RICH.
FT DOMAIN 1267 1302 GLU-RICH.
FT DOMAIN 1411 1414 POLY-GLU.
FT DOMAIN 1593 1597 POLY-SER.
FT DOMAIN 1643 1704 GLN/PRO-RICH.
FT DOMAIN 1897 1977 MET-RICH.
FT SITE 1546 1547 BREAKPOINT FOR TRANSLATION TO FORM MOZ-CBP.
SQ SEQUENCE 2004 AA; 225054 MW; 9FBBAC3792854BA CRC64;

Query Match 6.9%; Score 114.5; DB 1; Length 2004;
Best Local Similarity 23.7%; Pred. No. 3;
Matches 40; Conservative 25; Mismatches 73; Indels 31; Gaps 5;

OY 47 SHRTSGSGLAIGS-OVITDSQGVNRYQVQGDIV-----SKIAQRYGLMREIGHINN 100
DB 1542 SQQVVDSEFSDIGSIESTTENENPSYDSTWGSICGNSSQSCSTG-----GLSSSS 1597
OY 101 LNSSYTIYTGWLTLMGDLKVERSSISGVNTA-----HTSPYAV 142
DB 1598 LTQSSCVLTQGMASGSSCSMMQSSVQPANCSIKSPQSCVERPSPNQQPPPPPPQ 1657
OY 143 QSSRPVQOHPAVQKPTPPVVVVVKKPTPPPPVQOPAPVAPVTEAPPA 191
DB 1658 QQPPEPPQOPAPQPPPPQOQOPOPOPP---QQPPPPPPQOQPPPP 1703

RESULT 29
HESI_RAT STANDARD; PRT; 281 AA.
ID HESI_RAT
AC Q04666;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Transcription factor HES-1 (Hairy and enhancer of split 1) (Hairy-like) (RH1).
GN HES1 OR HES-1 OR HL.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxId=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Embryo;
RX MEDLINE=94040724; PubMed=1340473;
RA Sasaki Y., Kagiyama R., Tagawa Y., Shigemoto R., Nakanishi S.;
RT "Two mammalian helix-loop-helix factors structurally related to Drosophila hairy and enhancer of split.";
RL Genes Dev. 6:2620-2634(1992).
[2]
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93109293; PubMed=8417318;
RA Feder J.N., Jan L.Y., Jan Y.-N.;
RT "A rat gene with sequence homology to the Drosophila gene hairy is rapidly induced by growth factors known to influence neuronal

```

```

RT differentiation.";
RL Mol. Cell. Biol. 13:105-113(1993).
CC -1- FUNCTION: TRANSCRIPTIONAL REPRESSOR OF GENES THAT REQUIRE A BHLH PROTEIN FOR THEIR TRANSCRIPTION. MAY ACT AS A NEGATIVE REGULATOR OF MYOGENESIS BY INHIBITING THE FUNCTIONS OF MYOD AND AHL1.
CC -1- SUBUNIT: TRANSCRIPTION REPRESSION REQUIRES FORMATION OF A COMPLEX WITH A CO-REPRESSOR PROTEIN (GROUCHO).
CC -1- TISSUE SPECIFICITY: PRESENT IN ALL TISSUES EXAMINED BUT HIGHEST IN EPITHELIAL CELLS AND IN MESODERM-DERIVED TISSUES SUCH AS EMBRYONAL MUSCLE CELLS.
CC -1- DOMAIN: HAS A PARTICULAR TYPE OF BASIC DOMAIN (PRESENCE OF A HELIX-INTERRUPTING PROLINE) THAT BINDS TO THE N-BOX (CAONAG), RATHER THAN THE CANONICAL E-BOX (CANNTG).
CC -1- DOMAIN: THE CARBOXYL-TERMINAL WRPW MOTIF IS A TRANSCRIPTIONAL REPRESSION DOMAIN NECESSARY FOR THE INTERACTION WITH GROUCHO, A TRANSCRIPTIONAL CO-REPRESSOR RECRUITED TO SPECIFIC TARGET DNA BY HAIRY-RELATED PROTEINS.
CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF TRANSCRIPTION FACTORS. "HAIRY" SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.ibt-sib.ch/announce/ or send an email to license@ibt-sib.ch).
CC -----
CC EMBL; D13417; BA02682.1; -.
CC EMBL; L04527; AAA41307.1; -.
CC PIR; S36748; S36748.
CC TRANSFAC; T01648; -.
CC InterPro; IPR001092; HLH_basic.
CC InterPro; IPR001650; Orange.
DR Pfam; PF00010; HLH; 1.
DR SMART; SMO0353; HLH; 1.
DR SMART; SMO0511; ORANGE; 1.
DR PROSITE; PS00038; HLH_1; 1.
DR PROSITE; PS50888; HLH_2; 1.
KW Transcription regulation; DNA-binding; Nuclear protein; Repressor.
FT DNA_BIND 35 47 BASIC DOMAIN.
FT DOMAIN 48 92 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
FT DOMAIN 156 247 PRO-RICH.
FT DOMAIN 250 274 SER/THR-RICH.
FT DOMAIN 276 279 WRPW MOTIF (REQUIRED FOR ACTIVITY) (BY SIMILARITY).
FT CONFLICT 86 86 K->N (IN REF. 2).
FT CONFLICT 93 93 R->W (IN REF. 2).
SQ SEQUENCE 281 AA; 29622 MW; 8A98C9946F075B0D CRC64;

Query Match 6.8%; Score 114; DB 1; Length 281;
Best Local Similarity 26.9%; Pred. No. 0.36;
Matches 42; Conservative 16; Mismatches 42; Indels 56; Gaps 8;

OY 126 SISSGVNT-----AHTSPVA-VQSSRPVQOHPAVQKPTPPVVVVVKKPT----- 169
DB 126 STCGSVNTEVTRILGHILANCTOINAMTYTGQHPALQADPPP-----PSGGGPGPOHA 180
OY 170 ---FTPPVQOPAPVAPVTEAPPATGS-----SGVMQFRYPVGA-- 206
DB 181 PFAPPPPLVPIPGGAAPPSPGAPCKLQSGQAGEAKVFGQVVPAPDQGFALIPNGAFA 240
OY 207 -TNVYKRF---GTAT-----VAGSTVTSNGMW 230
DB 241 HSGVPIPVYTSNGTSVGPNAVSPSSGSLTADSMW 276

RESULT 30
MANB_CALSA STANDARD; PRT; 1331 AA.
ID MANB_CALSA
AC P22533;
DT 01-AUG-1991 (Rel. 19, Created)

```



```

CC -1- DOMAIN: HAS A PARTICULAR TYPE OF BASIC DOMAIN (PRESENCE OF A
CC HELIX-INTERPRETING PROLINE) THAT BINDS TO THE N-BOX (CACNAG),
CC RATHER THAN THE CANONICAL E-BOX (CANTG).
CC -1- DOMAIN: THE CARBOXYL-TERMINAL WRPW MOTIF IS A TRANSCRIPTIONAL
CC REPRESSION DOMAIN NECESSARY FOR THE INTERACTION WITH GROUCHO, A
CC TRANSCRIPTIONAL CO-REPRESSOR RECRUITED TO SPECIFIC TARGET DNA BY
CC HAIRY-RELATED PROTEINS.
CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC TRANSCRIPTION FACTORS. "HAIRY" SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL, L19314; AAA65220.1; -.
CC EMBL, AF264785; AAF73060.1; -.
CC EMBL, AK000415; BAA91149.1; -.
CC TRANSFAC; T04892; -.
CC Genew; HGNC:5192; HRX.
CC MIM; 139605; -.
CC InterPro; IPR001092; HLH_basic.
CC InterPro; IPR003650; Orange.
CC Pfam; PF00010; HLH; 1.
CC SMART; SM00353; HLH; 1.
CC SMART; SM00511; ORANGE; 1.
CC PROSITE; PS00038; HLH_1; 1.
CC PROSITE; PS00888; HLH_2; 1.
CC Transcription regulation; DNA-binding; Nuclear protein; Repressor.
CC Transcription regulation; DNA-binding; Nuclear protein; Repressor.
CC FT DNA BIND 35 47 BASIC DOMAIN.
CC FT DOMAIN 48 92 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
CC FT DOMAIN 156 246 PRO-RICH.
CC FT DOMAIN 249 273 SER/THR-RICH.
CC FT DOMAIN 275 278 WRPW MOTIF (REQUIRED FOR ACTIVITY)
CC (BY SIMILARITY).
CC SQ SEQUENCE 280 AA; 29541 MW; F9342A88FC749E3C CRC64;

Query Match 6.6%; Score 109.5; DB 1; Length 280;
Best Local Similarity 27.3%; Pred. No. 0.73;
Matches 41; Conservative 15; Mismatches 49; Indels 45; Gaps 7;

QY 126 SISSGVNT-----AHTSPVA-VQSSRPVQOHAPVOKPPVAVVKKP-----TTP 172
DB 126 STCEVNTVEVTRRLGLHLANCMTQINAMTYEQPHALQAPPFPFGGPGPHAFAPPP 185
QY 173 PVVQOPAVAPVTEAPPATGS-----SGVMQFRYPVGA--TNPV 211
DB 186 PLVPFPGAAPPPGAPFCGLSGAGEAAKVFQGVVPVAPDQFAFLIPNGAFHSGFVI 245
QY 212 RRF-----GTAT-----VAGSTVTSNGMW 230
DB 246 PVTYSNGTISVGNAPNAVSPSSGSLTADSMW 275

RESULT 32
MYSC_DICDI STANDARD; PRT; 1181 AA.
AC P42522;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin IC heavy chain.
GN MYOC OR DMIC.
OS Dictyostelium discoideum (slime mold).
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
OX NCBI_Taxid=44689;
RX STRAIN=AX2;
MEDLINE=95348228; PubMed=7622596;

```

```

RA Peterson M.D., Novak K.D., Reedy M.C., Ruman J.I., Titus M.A.;
RT "Molecular genetic analysis of myoc, a Dictyostelium myosin I.";
RL J. Cell Sci. 108:1093-1103(1995).
CC -1- FUNCTION: MYOSIN IS A PROTEIN THAT BINDS TO ACTIN & HAS ATPASE
CC ACTIVITY THAT IS ACTIVATED BY ACTIN.
CC -1- SUBUNIT: MYOSIN I HEAVY CHAIN IS SINGLE-HEADED. DIMER OF A HEAVY
CC AND A LIGHT CHAIN. INABILITY TO SELF-ASSEMBLE INTO FILAMENTS.
CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL, L35323; AAC37427.1; -.
CC HSSP; P08799; IMND.
CC Dictydb; DD01090; myoc.
CC InterPro; IPR001452; SH3.
CC InterPro; IPR001609; myosin_head.
CC Pfam; PF00018; SH3; 1.
CC Pfam; PF00063; myosin_head; 2.
CC PRINTS; PR00193; MYOSINHEAVY.
CC PRINTS; PR00452; SH3DOMAIN.
CC Prodom; PD000066; SH3; 1.
CC Prodom; PD000355; myosin_head; 1.
CC SMART; SM00242; MYSC; 1.
CC SMART; SM00326; SH3; 1.
CC PROSITE; PS50002; SH3; 1.
CC PROSITE; PS50088; SH3; 1.
CC Myosin; Actin-binding; ATP-binding; Chemotaxis; SH3 domain;
CC Multigene family.
CC FT DOMAIN 1 2 MYOSIN HEAD-LIKE.
CC FT NP BIND 109 116 NON ALPHA-HELICAL, C-TERMINAL DOMAIN.
CC FT DOMAIN 1122 1181 ATP (POTENTIAL).
CC SQ SEQUENCE 1181 AA; 132915 MW; 5EB1EE47F0CA8E03 CRC64;

Query Match 6.6%; Score 109.5; DB 1; Length 1181;
Best Local Similarity 35.0%; Pred. No. 3.6;
Matches 36; Conservative 17; Mismatches 31; Indels 19; Gaps 6;

QY 120 LKVERSISSGVNTAHT-----PSPVAVQS--RPVQOHAPVOKTPTVVVKKRTP 170
DB 943 LKKQOFHIASGLPASTYVAKVRKNSOVSTPSKFIAPVAK-PVAKRSGSVIMKKPAP 1001
QY 171 T-----PVVQOPAVAPV--PVTAPF-ATGSSGVNQFRYPV 204
DB 1002 AAPSGPPVMKKPAPFAPGAPMMKKAPAPGAPMMKKPAPV 1044

RESULT 33
VGP3_EBV STANDARD; PRT; 907 AA.
AC P03200; P03201;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Envelope glycoprotein GP340 (Membrane antigen) (MA) [Contains:
DE Glycoprotein GP220].
GN BLF1.
OS Epstein-barr virus (strain B95-8) (human herpesvirus 4).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Lymphocryptovirus.
OX NCBI_Taxid=10377;
RX MEDLINE=84270667; PubMed=6087149;
RA Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,
RA Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seguin C.,
RA Tuffnell P.S., Barrell B.G.;

```

"DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";
 Nature 310:207-211 (1984).
 CC - FUNCTION: RESPONSIBLE FOR EBV BINDING TO THE CR2 RECEPTOR ON HUMAN
 CC B-CELLS.
 CC - SUBCELLULAR LOCATION: MOST ABUNDANT COMPONENT OF THE VIRAL
 CC ENVELOPE.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: V01555; CAA24854.1; -
 DR PIR: A03762; Q0BE21.
 DR PIR: A03763; S33008.
 DR PIR: S33008; S33008.
 KW Membrane; Glycoprotein; Antigen; Late protein; Alternative splicing.
 FT CARBOHYD 47 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 87 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 114 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 166 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 169 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 195 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 229 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 277 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 318 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 328 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 345 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 356 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 378 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 386 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 411 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 435 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 443 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 457 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 487 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 519 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 533 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 547 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 568 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 589 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 610 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 624 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 627 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 656 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 683 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 701 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 735 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 746 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 755 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 780 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 815 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 858 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 888 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARSPLIC 502 MISSING (IN GP220).
 SQ SEQUENCE 907 AA; 94431 MW; 0750141CBGAC5C9 CRC64;

Query Match
 Best Local Similarity 6.5%; Score 109; DB 1; Length 907;
 Matches 66; Conservative 34; Mismatches 122; Indels 62; Gaps 13;

QY 35 SKPTVSTSGSGSHRISGGLAIGSOVITDSOGVPRRY--QVKQDPTSKIAKQRIGLAW 92
 DB 430 TSPTLTGTFADNPTTIG-----LPSSTHVPNLTPASTGPTVS----- 469
 QY 93 REIGHNININSSYTIYTGMLT---LWSGLDKVERRSISGCVNATHTSPVAVOSSRPP 148
 DB 470 --TADVTSPTPAGTSGASPVTPSPSPWONGTESKAPDMTSTSPVTTPTPNATSPPT--- 524

QY 149 VOQHAPVOKPPTP-----FVVVKKPTP--TPVVQOPAPVAPVTEAPATG-----SSGV 197
 DB 525 ----PATTPTPNATSPTPATPTPNATSPTLKTSPTSAVTPPNATSPTLKTSPT 580
 QY 198 MQFRYPV-GATNPVRRFGTAVAGSTVTSNGMWF--GRDDLLINAGNACTVIOADHMD 255
 DB 581 SAVTTPPNATSPTLGK-TSPTSAVTPPTPNATPTVGETSPPOANATN-----HTLG 631
 QY 256 GASIVTGHNTGTFSSYHIKDAQVKTGDTVATGARIAMKNQPS 299
 DB 632 GTS-----PTPVTSOPKNATSAVTTGCHNTTSSSTSMISRPS 670

RESULT 34
 ID EXTN SORBI STANDARD; PRT; 283 AA.
 AC P24152;
 DI 01-MAR-1992 (Rel. 21, Created)
 DI 01-MAR-1992 (Rel. 21, Last sequence update)
 DI 30-MAR-2000 (Rel. 39, Last annotation update)
 DE Extensin precursor (Proline-rich glycoprotein).
 GN HRGP.
 OS Sorghum bicolor (Sorghum) (Sorghum vulgare).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
 OC Panicoideae; Andropogoneae; Sorghum.
 OC NCBI_TaxID=4558;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Leaf;
 RX MEDLINE=91370882; PubMed=1893107;
 RA Raz R., Cretin C., Puigdomenech P., Martinez-Izquierdo J.A.;
 RT "The sequence of a hydroxyproline-rich glycoprotein gene from sorghum
 vulgare.";
 RT Plant Mol. Biol. 16:365-367(1991).
 CC - FUNCTION: STRUCTURAL COMPONENT IN PRIMARY CELL WALL.
 CC - PTM: EXTENSIN CONTAIN A CHARACTERISTIC REPEAT OF THE PENTAPEPTIDE
 CC SER-PRO(4). THE PROLINE RESIDUE IS HYDROXYLATED AND THEN
 CC GLYCOSYLATED.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X56010; CAA39485.1; -
 DR PIR: S14449; S14449.
 DR Interpro; IPR002965; P-rich extensin.
 DR PRINTS; PR01217; PRICEXTENSIN.
 KW Repeat; Cell wall; Glycoprotein; Signal; Structural protein;
 KW Hydroxylation.
 FT SIGNAL 1 24 POTENTIAL.
 FT CHAIN 25 283 EXTENSIN.
 SQ SEQUENCE 283 AA; 29593 MW; 8D7FCDD0D8ED2D90 CRC64;

Query Match
 Best Local Similarity 6.5%; Score 108.5; DB 1; Length 283;
 Matches 27; Conservative 7; Mismatches 23; Indels 15; Gaps 3;

QY 135 HRPSPVAVOSSRP--PVOQHAPVOKPPTPVPVVVKKPTPVPVQOPAP-----VA 182
 DB 117 YTPSPKPAKTKPPTPTPKPAKTKPPTPVVTV--PSKRPVTKPPTPKPPTPVVTPMVK 173
 QY 183 PPTVTEAPATGS 194
 DB 174 PPTVTKPPTHTPS 185

RESULT 35
 GPL_CHURE

```

ID GPI CHLRE STANDARD; PRT; 555 AA.
AC Q9FP06; Q03927;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Vegetative cell wall protein gpi precursor (Hydroxyproline-rich
DE glycoprotein 1).
GN GPI.
OS Chlamydomonas reinhardtii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadaceae; Chlamydomonas.
OX NCBI_TaxID=3055;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21159092; PubMed=11258910;
RA Perrie P.J., Moesener J.P., Waffenschmidt S., Kitz S., Drees J.,
RA Goodenough U.W.;
RT "Glycosylated polypeptide II rods-with-kinks as a structural motif in
RT plant hydroxyproline-rich glycoproteins.";
RL Biochemistry 40:2978-2987(2001).
RN [2]
RP PARTIAL PRELIMINARY SEQUENCE FROM N.A.
RX MEDLINE=91017504; PubMed=1699225;
RA Adair W.S., Apt K.E.;
RT "Cell wall regeneration in Chlamydomonas: accumulation of mRNAs
RT encoding cell wall hydroxyproline-rich glycoproteins.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:7355-7359(1990).
CC -1- FUNCTION: Major component of the outer cell wall W6 (crystalline)
CC layer.
CC -1- SUBUNIT: Associates with GP2 and GP3.
CC -1- PTM: N-glycosylated and O-glycosylated.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib.ch).
CC -----
CC EMBL; AF309494; AAG45420.1; -.
CC DR EMBL; M58496; AAG49706.1; ALT_SEQ.
CC DR GlycoSuiteDB; Q9FP06; -.
CC DR InterPro; IPR002965; P-rich_extensin.
CC DR InterPro; IPR003882; Pichl1_extensin.
CC DR PRINTS; PR01217; PRICHEXTENSIN.
CC DR PRINTS; PR01218; PSTLXTENSIN.
CC DR Glycoprotein; Repeat; Signal.
CC FT SIGNAL 1 29
CC FT CHAIN 30 555 VEGETATIVE CELL WALL PROTEIN GPI.
CC FT DOMAIN 40 339 49 X 5 AA APPROXIMATE PSEXP REPEATS.
CC FT DOMAIN 259 279 POLY-PRO.
CC FT CARBOHYD 359 399 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 455 455 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 493 493 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC SEQUENCE 555 AA; 54219 MW; 6A584A9046502F5 CRC64;
CC -----
Query Match 6.5%; Score 108.5; DB 1; Length 555;
Best Local Similarity 44.6%; Pred. No. 1.8;
Matches 25; Conservative 4; Mismatches 18; Indels 9; Gaps 2;

```

```

136 TPSPVAVOSRRPPVQOHPAVVKKPTPPVVOQAP--VAPVTEAP 189
DB 154 SPSPVPSPSPSPVPPSPAPSPPTP-----PSPSVPSPSPAPSPVPPSP 202

```

```

RESULT 36
ID F12255;
AC P12255;
ID F12255; STANDARD; PRT; 3591 AA.
AC P12255;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)

```

```

DE Filamentous hemagglutinin.
GN FHAB.
OS Bordetella pertussis.
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC Bordetella.
OX NCBI_TaxID=520;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=90355839; PubMed=2388559;
RA Reiman D.A., Domenighini M., Tuomanen E., Rappuoli R., Falkow S.;
RT "Genetic characterization of Bordetella pertussis filamentous
RT haemagglutinin: a protein processed from an unusually large
RT precursor.";
RL Mol. Microbiol. 4:787-800(1990).
RN [2]
RP SEQUENCE OF 1-3261 FROM N.A.
RX MEDLINE=8920384; PubMed=239596;
RA Reiman D.A., Domenighini M., Tuomanen E., Rappuoli R., Falkow S.;
RT "Filamentous hemagglutinin of Bordetella pertussis: nucleotide
RT sequence and crucial role in adherence.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:2637-2641(1989).
CC -1- FUNCTION: EVIDENCE FOR A ROLE IN HOST-CELL BINDING AND
CC INFECTON.
CC -1- SUBCELLULAR LOCATION: SURFACE.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib.ch).
CC -----
CC EMBL; M60351; AAA22874.1; -.
CC DR EMBL; M60351; AAA22875.1; ALT_INIT.
CC DR EMBL; M60351; AAA22976.1; ALT_INIT.
CC KW Antigen; Hemagglutinin.
CC SEQUENCE 3591 AA; 367420 MW; EF7418B30D6E5138 CRC64;
CC -----
Query Match 6.5%; Score 108.5; DB 1; Length 3591;
Best Local Similarity 24.5%; Pred. No. 15;
Matches 54; Conservative 21; Mismatches 74; Indels 71; Gaps 11;

```

```

6 AINSONOKPIRLGIVITTCILAGCASKPTYNSTSGSHRSGGLAIG--SQVI 63
DB 3165 AVNAQNLK-----DYRDKG-----GGGNAVIGSITTL 3193
QY 64 TDSQGV-----NRQYKQGDV-----SKIAQRYGLMREIGHINNINSSVTIYTGWL 113
DB 3194 APTVGVAFGRYVAGBDYQAEQATIDVGQTKDPAFLQGVGKGLTNDQAQATV--QRN 3251
QY 114 TLNKG---DLKVRERST--SSGVNTATPPSPVAVOS--SRP-----PVOQHNAVQKP 158
DB 3252 KHMAGGSGSEFSAKSKKKQVPRVETPTDVAVDGPPSRPTTPPASPPIRATVEVSSP 3311
QY 159 TP-----PVVVKKPTPPVVOQHPAVVKKPTPPVVOQAPVAPVTE 187
DB 3312 PPVSVATVEVPRKRVETAPLPPRPVAAQVVPVTPPKVE 3351

```

```

RESULT 37
ID A112 HUMAN
AC P56397; STANDARD; PRT; 1593 AA.
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE ADAMTS-12 precursor (EC 3.4.24.-) (A disintegrin and
DE metalloproteinase with thrombospondin motifs 12) (ADAM-TS 12).
GN ADAMTS12.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

OC Mammalia; Euthera; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC TISSUE=Fetal Lung;
 RA MEDLINE=21264577; PubMed=11279086;
 RX Cal S.; Argueelles J.M.; Fernandez P.L.; Lopez-Otin C.;
 RT "Identification, characterization, and intracellular processing of
 RT ADAM-12, a novel human disintegrin with a complex structural
 RT organization involving multiple thrombospondin-1 repeats.";
 RL J. Biol. Chem. 276:17933-17940(2001).
 CC -1- COPACITOR: BINDS 1 ZINC ION (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
 CC matrix (By similarity).
 CC -1- TISSUE SPECIFICITY: Expressed exclusively in fetal lung. Is widely
 CC expressed in gastric carcinomas and in cancer cells of diverse
 CC origin.
 CC -1- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE 1 DOMAINS ARE IMPORTANT
 CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX (BY
 CC SIMILARITY).
 CC -1- PM: THE PRECURSOR IS CLEAVED BY A PURIN ENDOPEPTIDASE.
 CC -1- PM: IS SUBJECTED TO AN INTRACELLULAR MATURATION PROCESS LEADING
 CC TO A FRAGMENT CONTAINING THE N-TERMINAL REGION INCLUDING THE
 CC METALLOPROTEINASE, DISINTEGRIN-LIKE, CYS-RICH AND TS-1 DOMAINS AND
 CC THE C-TERMINAL FRAGMENT CONTAINING THE SPACER 2 AND THE FOUR TS-1
 CC DOMAINS.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B.
 CC -1- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 8 TSP TYPE-1 DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AJ250725; CAC20419.1; -
 CC Genew; HGNC:14605; ADAMTS12.
 DR MIM; 606184; -
 DR InterPro; IPR001762; Disintegrin.
 DR InterPro; IPR002870; Pep_M12B_propep.
 DR InterPro; IPR001590; Repolysin.
 DR InterPro; IPR000884; TSP1.
 DR InterPro; IPR000130; Zn_MTpeptidase.
 DR Pfam; PF00090; TSP_1; 6.
 DR Pfam; PF01421; Repolysin; 1.
 DR Pfam; PF01562; Pep_M12B_propep; 1.
 DR SMART; SM00209; TSP1; 8.
 DR PROSITE; PS50215; ADAM_MEPRO; 1.
 DR PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
 DR PROSITE; PS00092; TSP1; 2.
 DR PROSITE; PS00142; ZINC_PROTEASE; 1.
 DR HydroLase; Metailprotease; Zinc; signal; Glycoprotein; Zymogen;
 KM Repeat; Extracellular matrix. POTENTIAL.
 FT SIGNAL 1 25
 FT PROPEP 26 240
 FT CHAIN 241 1393
 FT DOMAIN 465 544
 FT DOMAIN 545 596
 FT DOMAIN 597 700
 FT DOMAIN 701 826
 FT DOMAIN 827 881
 FT DOMAIN 886 943
 FT DOMAIN 947 995
 FT DOMAIN 996 1315
 FT DOMAIN 1316 1364
 FT DOMAIN 1367 1423
 FT DOMAIN 1426 1471
 FT DOMAIN 1471 1471
 FT DOMAIN 1471 1471
 FT DOMAIN 1471 1471
 FT SITE 208 208
 CYS-TEINE SWITCH (POTENTIAL).

FT METAL 392 392 ZINC (CATALYTIC) (BY SIMILARITY).
 FT ACT_SITE 393 393 BY SIMILARITY.
 FT METAL 396 396 ZINC (CATALYTIC) (BY SIMILARITY).
 FT METAL 402 402 ZINC (CATALYTIC) (BY SIMILARITY).
 FT CARBOHYD 105 105 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 125 125 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 215 215 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 485 485 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 685 685 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 790 790 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 951 951 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1104 1104 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1275 1275 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1300 1300 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1320 1320 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1371 1371 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1378 1378 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1503 1503 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 1593 AA; 177545 MW; 07F9F48E3BD83A3 CRC64;
 Query Match 6.5%; Score 108; DB 1; Length 1593;
 Best Local Similarity 21.2%; Pred. No. 6.4;
 Matches 57; Conservative 35; Mismatches 93; Indels 84; Gaps 10;
 QY 80 TVSKIAQRYGLNWEIGHINNLSYTYTGQW---LTLWSGDLKRYERSIS----- 129
 DB 1054 TASKEGDGLGKQWQDSSTOPELSSRYLISTGSTSQPILTSGLSIOPSEEVSSSDTCPT 1113
 QY 130 --GNVTAHPSPVAVOSRPPV-----OCHPAVOKPTPP 161
 DB 1114 SEGGLVATTTGSGGLSSSRNITWVPFVNTLTKGPEMEIHSNGEREQPEPKDESNP 1173
 QY 162 VV--VVKPTPTPPVQCAPVAPVPE-----APFATSSGVQWQFRPVGATNPV 211
 DB 1174 VIVTKIRFGNDAPFESTEMFLAPPLTDLRSRWPFPSVYMGDLPE-----SQRTT 1227
 QY 212 RRFQATVATGATVTSNGMFWGSGRQDILNANACTVIO--ADHMDGASIVYQHTNGPVS 269
 DB 1228 SETGPRFEG---WTEKPRANTLPLUGDHPQ-----PSGKTA 1263
 QY 270 SYTHIK---DAVKTGDTVRTGQRIASM 294
 DB 1264 NNNHLKLPNNNQTSSEPVLTEDATSL 1292
 RESULT 38
 ID ALYS_ENTFA STANDARD; PRT; 671 AA.
 AC P37710;
 DT 01-OCT-1994 (Rel. 30; Created)
 DT 01-OCT-1994 (Rel. 30; Last sequence update)
 DT 16-OCT-2001 (Rel. 40; Last annotation update)
 DE Autolysin precursor (EC 3.2.1.-) (Peptidoglycan hydrolase)
 DB (Bacteria: glycosidase).
 OS Enterococcus faecalis (Streptococcus faecalis).
 OC Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.
 OX NCBI_TaxID=1351;
 RN (1)
 RP SEQUENCE FROM N.A.
 RA MEDLINE=91358349; PubMed=1679432;
 RA Bellevue C.; Potvin C.; Trudel J.; Asselin A.; Bellemare G.;
 RT "Cloning, sequencing, and expression in Escherichia coli of a
 RT Streptococcus faecalis autolysin.";
 RL J. Bacteriol. 173:5619-5623(1991).
 CC -1- FUNCTION: HYDROLYZES THE CELL WALL OF E. FAECALIS AND
 CC M. LYSOBIOTICUS. MAY PLAY AN IMPORTANT ROLE IN CELL WALL GROWTH
 CC AND CELL SEPARATION.
 CC -1- SUBCELLULAR LOCATION: Secreted (Probable).
 CC -1- DOMAIN: LYSM REPEATS ARE THOUGHT TO BE INVOLVED IN PEPTIDOGLYCAN
 CC BINDING.
 CC -1- SIMILARITY: BELONGS TO FAMILY 73 OF GLYCOSYL HYDROLASES.
 CC -1- SIMILARITY: CONTAINS 5 LYSM REPEATS.
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; M58002; AAA67325.1; -
 DR PIR; A38109; A38109.
 DR InterPro; IPR002901; Amidase_4.
 DR InterPro; IPR002482; LYSM.
 DR Pfam; PF01476; LYSM; 5.
 DR Pfam; PF01832; Amidase_4; 1.
 DR SMART; SMO0047; LY22; 1.
 DR SMART; SMO0257; LYSM; 5.
 DR Hydrolyase; Glycosidase; Bacteriolytic enzyme; Cell wall;
 KM Cell division; Septation; Repeat; Signal.
 FT SIGNAL
 FT CHAIN 1 53
 FT REPEAT 54 671
 FT REPEAT 363 405
 FT REPEAT 431 473
 FT REPEAT 499 541
 FT REPEAT 567 609
 FT REPEAT 629 671
 SQ SEQUENCE 671 AA; 70442 MW; 34582686C6C1A4A33 CRC64;

Query Match 6.4%; Score 107.5; DB 1; Length 671;
 Best Local Similarity 20.8%; Pred. No. 2.6;
 Matches 66; Conservative 43; Mismatches 77; Indels 131; Gaps 15;

QY 34 ASKPTVNS-----TSGSGSHRTSGSGGALISQVITTDGQVNRVYQVQG 78
 DB 311 ATDPSYNAKNNVITAYNLGYDITPSSGNN--TGGCTVNPETGSGNNSGNTTYTVVSG 368
 QY 79 DTVSKIAQRVGL-----NMREI-----GHINNLSY 105
 DB 369 DTLNKIAQYGVSVANLRSNMGISGDLIFVGQKLIYKKGASGNTGSGNNSGNTN 428
 QY 106 TTYGQMTLWMSGDLKVERSSGVNTHATPSPVAVOSSRPVQGHAVQKPTPVVVV 165
 DB 429 TTY-----TVKSGD--TLNKIAQYGVTVANLRSNMGISGDLIFVGQK-----LIV 472
 QY 166 KKPFTPVVQGPAPVAPVTEAPATGSSGVNQPRVPGATNPVRRFGATVAGSTVT 225
 DB 473 KKGFT-----SGNTG-----GSSNN 491
 QY 226 SNG--MWSGRDGLINASNGTVIQADNMDSGIVIQHTNGFVSSYHIKDAQVKTGD 283
 DB 492 QSGNTVYTVTIKSGDTLTK-----IAQYGVSVAN--LRSNMGI-----SGD 530
 QY 284 TVRTGQRIASMKNQPSG 300
 DB 531 LIFAGQKLI-VKGTSG 546

RESULT 39
 PRZE_MYXA
 ID PRZE_MYXA STANDARD; PRT; 777 AA.
 AC P18769;
 DT 01-NOV-1990 (Rel. 16; Created)
 DT 01-NOV-1990 (Rel. 16; Last sequence update)
 DT 15-JUN-2002 (Rel. 41; Last annotation update)
 DE Gliding motility regulatory protein (EC 2.7.3.-).
 GN PRZE.
 OS Myxococcus xanthus.
 OC Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
 OC Myxococcales; Cyetobacterineae; Myxococcaceae; Myxococcus.
 OX NCBI_TaxID=34;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90332690; PubMed=2165608;
 RA McCleary W.R.; Zusman D.R.;

RT "Frze of Myxococcus xanthus is homologous to both Chea and Chey of
 RT Salmonella typhimurium."
 RL Proc. Natl. Acad. Sci. U.S.A. 87:5898-5902(1990).
 RN [2]
 RP PHOSPHORYLATION OF HIS-49.
 RX MEDLINE=91072208; PubMed=2123853;
 RA McCleary W.R.; Zusman D.R.;
 RT Purification and characterization of the Myxococcus xanthus Frze
 RT protein shows that it has autophosphorylation activity."
 RL J. Bacteriol. 172:6661-6668(1990).
 CC - FUNCTION: FRZE IS INVOLVED IN A SENSORY TRANSDUCTION PATHWAY THAT
 CC CONTROLS THE FREQUENCY AT WHICH CELLS REVERSE THEIR GLIDING
 CC DIRECTION. FRZE SEEMS TO BE CAPABLE OF AUTOPOSPHORYLATING ITSELF
 CC ON AN HISTIDINE RESIDUE AND THEN TO TRANSFER THAT GROUP TO AN
 CC ASPARTATE RESIDUE IN THE C-TERMINAL PART OF THE PROTEIN.
 CC - SIMILARITY: CONTAINS 1 HISTIDINE KINASE DOMAIN.
 CC - SIMILARITY: SIMILAR TO BOTH CHEA AND CHEY.
 CC - SIMILARITY: CONTAINS 1 RESPONSE REGULATORY DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; M35192; AAA25396.1; -
 DR PIR; A35966; A35966.
 DR HSSP; Q56310; 1B3Q.
 DR InterPro; IPR003594; ATPbind_Atpase.
 DR InterPro; IPR004358; Bact_sense_pr_C.
 DR InterPro; IPR002545; Chew_1.
 DR InterPro; IPR004359; HIS_KIN_sig.
 DR InterPro; IPR002570; Hpt.
 DR InterPro; IPR001789; Response_reg.
 DR Pfam; PF00072; response_reg; 1.
 DR Pfam; PF01584; Chew; 1.
 DR Pfam; PF01627; Hpt; 1.
 DR Pfam; PF02518; HATPase_C; 1.
 DR PRINTS; PR00344; BCTRLENSOR.
 DR ProDom; PD000039; Response_reg; 1.
 DR ProDom; PD003142; Hpt; 1.
 DR SMART; SMO0260; Chew; 1.
 DR SMART; SMO0387; HATPase_C; 1.
 DR SMART; SMO0073; HPT; 1.
 DR SMART; SMO0448; REC; 1.
 DR PROSITE; PS50109; HIS_KIN; 1.
 DR PROSITE; PS50110; RESPONSE_REGULATORY; 1.
 KW Sensory transduction; Transferrase; Kinase; Phosphorylation.
 FT DOMAIN 270 509
 FT DOMAIN 660 776
 FT MOD RES 49 49
 FT DOMAIN 130 197
 SQ SEQUENCE 777 AA; 83189 MW; 9912BD40991C69E5 CRC64;

Query Match 6.4%; Score 107.5; DB 1; Length 777;
 Best Local Similarity 23.7%; Pred. No. 3.1;
 Matches 53; Conservative 30; Mismatches 76; Indels 65; Gaps 11;

QY 110 GQMTLWMSGDLKVERSSISGVNT-----AHPSPVA-----VQSRPP--VQGHFA 154
 DB 80 GKRTVATVTLVLAACVLSLNDLDSGANTGNASEMVMALAEVSQGTPTALAGARPV 139
 QY 155 VQKPTPVVVVKKPTFPVPPVQGP-----APVAPVTEAPPA--TGSSGVNQPRVPGATN 208
 DB 140 APPAPAPPAVPAVAVVTPPAVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPV 197
 QY 209 PVVRRFGATVAGSTVTSNGMWSFGRDGLINASNGTVIQADNMDSGIVIQHTNGFV 268
 DB 198 -----HGRDEAPSAKSAV--ADRSI--RVNVEVDALGLL 230
 QY 269 SSYIHXDAQVKTGDTVTRGRIASMKNQPSGALFFFRISRRG 312

```

Db          231 AGDLIVESAR-----GRUNSSERE-----ALFE-RFRLG 259

RESULT 40
SHKL_RAT    STANDARD; PRT; 2167 AA.
ID          SHKL_RAT
AC          Q9WV48; Q9WU13; Q9WU8; Q9QZ28;
DT          15-JUN-2002 (Rel. 41, Created)
DT          15-JUN-2002 (Rel. 41, Last sequence update)
DT          15-JUN-2002 (Rel. 41, Last annotation update)
DE          SH3 and multiple ankyrin repeat domains protein 1 (Shank1) (GKAP/SAPAP
DE          interacting protein) (SPANK-1) (Synamon) (Somatostatin receptor
DE          interacting protein) (SSTR interacting protein) (SSTRIP).
GN          SHANK1.
OS          Rattus norvegicus (Rat).
OC          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC          Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX          NCBI_TaxId=10116;
[1]
RP          SEQUENCE FROM N.A. (ISOFORMS 1 AND 3), AND INTERACTION WITH DLGAP1 AND
RP          DLG4.
RC          TISSUE=Brain;
RC          MEDLINE=994119021; PubMed=10488079;
RC          Yao I., Hata Y., Hirono K., Desnuchi M., Ide N., Takeuchi M., Takai Y.;
RC          "Synamon, a novel neuronal protein interacting with synapse-associated
RC          protein 90/postsynaptic density-95-associated protein.",
RC          J. Biol. Chem. 274:27463-27466(1999).
[2]
RP          SEQUENCE FROM N.A. (ISOFORM 4), AND INTERACTION WITH DLGAP1.
RC          STRAIN=Sprague-Dawley;
RC          MEDLINE=99360850; PubMed=10433368;
RC          Naisbitt S., Kim E., Tu D.C., Xiao B., Sala C., Valtchanoff J.,
RC          Weisberg R.J., Worley P.F., Sheng M.;
RC          "Shank, a novel family of postsynaptic density proteins that binds to
RC          the NMDA receptor/PSD-95/GKAP complex and cortactin.",
RC          Neuron 23:569-582(1999).
[3]
RP          SEQUENCE FROM N.A. (ISOFORM 1).
RC          TISSUE=Brain;
RC          MEDLINE=20549637; PubMed=10958799;
RC          Tobben S., Suedhof T.C., Stahl B.;
RC          "The G protein-coupled receptor Ctl1 interacts directly with proteins
RC          of the shank family.",
RC          J. Biol. Chem. 275:36204-36210(2000).
[4]
RP          PARTIAL SEQUENCE FROM N.A. (ISOFORMS 1; 2; 3; 4 AND 5), AND
RP          DEVELOPMENTAL STAGE.
RC          TISSUE=Brain;
RC          MEDLINE=99436166; PubMed=10506216;
RC          Lim S., Naisbitt S., Yoon J., Hwang J.I., Suh P.G., Sheng M., Kim E.;
RC          "Characterization of the Shank family of synaptic proteins. Multiple
RC          genes, alternative splicing, and differential expression in brain and
RC          development.",
RC          J. Biol. Chem. 274:29510-29518(1999).
[5]
RP          PARTIAL SEQUENCE FROM N.A. (ISOFORM 4).
RC          TISSUE=Brain;
RC          MEDLINE=200202175; PubMed=10551867;
RC          Zitzler H., Hoernck H.-H., Baechner D., Richter D., Kreienkamp H.-T.;
RC          "Somatostatin receptor interacting protein defines a novel family of
RC          multidomain proteins present in human and rodent brain.",
RC          J. Biol. Chem. 274:32997-33001(1999).
[6]
RP          INTERACTION WITH HOMER-1, AND SUBCELLULAR LOCATION.
RC          PubMed=10433269;
RC          Tu J.C., Xiao B., Naisbitt S., Yuan J.P., Petralia R.S., Brakeman P.,
RC          Doan A., Aakalu V.K., Lianhan A.A., Sheng M., Worley P.F.;
RC          "Coupling of mGluR/Homer and PSD-95 complexes by the Shank family of
RC          postsynaptic density proteins.",
RC          Neuron 23:583-592(1999).
[7]
RP          INTERACTION WITH SPTAN1.
RX          PubMed=11509555;
RX          Bockers T.M., Mameza M.G., Kreutz M.R., Bockmann J., Weise C.,
RX          Buck F., Richter D., Gundelfinger E.D., Kreienkamp H.-T.;
RX          "Synaptic scaffolding proteins in rat brain. Ankyrin repeats of the
RX          multidomain Shank protein family interact with the cytoskeletal
RX          protein alpha-fodrin.",
RX          J. Biol. Chem. 276:40104-40112(2001).
[8]
RX          FUNCTION.
RX          PubMed=11499055;
RX          Sala C., Piench V., Wilson N.R., Passafium M., Liu G., Sheng M.;
RX          "Regulation of dendritic spine morphology and synaptic function by
RX          Shank and Homer.",
RX          Neuron 31:115-130(2001).
[9]
RX          REVIEW.
RX          PubMed=10806096;
RX          Sheng M., Kim E.;
RX          "The Shank family of scaffold proteins.",
RX          J. Cell Sci. 113:1851-1856(2000).
-1- FUNCTION: Seems to be an adapter protein in the postsynaptic
density (PSD) of excitatory synapses that interconnects receptors
of the postsynaptic membrane including NMDA-type and metabotropic
glutamate receptors, and the actin-based cytoskeleton. May play a
role in the structural and functional organization of the
dendritic spine and synaptic junction. Overexpression promotes
maturation of dendritic spines and the enlargement of spine heads
via its ability to recruit Homer to postsynaptic sites, and
enhances presynaptic function.
-1- SUBUNIT: May homomultimerize via its SAM domain. Interacts with
SPTAN1, Homer-1 and DLGAP1/GKAP. Is part of a complex with
DLG4/PSD-95 and DLGAP1/GKAP. Interacts with SSTR2 C-terminus via
the PDZ domain (by similarity).
-1- SUBCELLULAR LOCATION: Cytoplasmic; postsynaptic density of
neuronal cells. Colocalizes with alpha-latrotoxin receptor 1.
-1- ALTERNATIVE PRODUCTS: At least 5 isoforms; 1, (shown here); 2, 3,
4/4 and 5, are produced by alternative splicing.
-1- TISSUE SPECIFICITY: Expressed only in brain (neocortical cortex,
CA1 region hippocampus and molecular layer of cerebellum).
-1- DEVELOPMENTAL STAGE: Expression increases from low levels at birth
to high levels at 3-4 weeks before dropping slightly in adulthood.
Expressed in the cortex and the molecular layer of the cerebellum
at postnatal day 7. Isoform 2 expression does not change during
development of both cortex and cerebellum. Isoform 4 expression
decreases significantly during development of cortex but not
cerebellum.
-1- SIMILARITY: BELONGS TO THE SHANK FAMILY.
-1- SIMILARITY: CONTAINS 7 ANK REPEATS.
-1- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
-1- SIMILARITY: CONTAINS 1 SAM DOMAIN.
-1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
-----
CC          This SWISS-PROT entry is copyright. It is produced through a collaboration
CC          between the Swiss Institute of Bioinformatics and the EMBL outstation
CC          at the European Bioinformatics Institute. There are no restrictions on its
CC          use by non-profit institutions as long as its content is in no way
CC          modified and this statement is not removed. Usage by and for commercial
CC          entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC          or send an email to license@isb-sib.ch).
-----
CC          EMBL; AF102855; AAD04569.2; -
CC          EMBL; AF131951; AAD29417.1; ALT_INIT.
CC          EMBL; AF159046; AAD42975.1; -
CC          EMBL; AF141904; AAF02498.1; ALT_INIT.
CC          HSSP; P00519; IAB1.
CC          InterPro; IPR002110; ANK.
CC          InterPro; IPR001478; PDZ.
CC          InterPro; IPR001600; SAM.
CC          InterPro; IPR001452; SH3.
CC          Pfam; PF00023; ank; 7.
CC          Pfam; PF00595; PDZ; 1.
CC          Pfam; PF00536; SAM; 1.
CC          Pfam; PF00018; SH3; 1.

```

Job time : 25 secs

DR	ProdDom:	PD000066;	SH3, 1.
DR	SMART:	SM00248;	ANK, 3.
DR	SMART:	SM00228;	PDZ, 1.
DR	SMART:	SM00454;	SAM, 1.
DR	SMART:	SM00326;	SH3, 1.
DR	PROSITE:	PS50088;	ANK_REPEAT, 3.
DR	PROSITE:	PS50297;	ANK_REP_REGION, 1.
DR	PROSITE:	PS50106;	PDZ, 1.
DR	PROSITE:	PS50002;	SH3, 1.
DR	PROSITE:	PS50105;	SAM_DOMAIN, 1.
KM	ANK_repeat:	SH3 domain;	Repeat; Alternative splicing.
FT	REPEAT	195	210
FT	REPEAT	212	245
FT	REPEAT	246	278
FT	REPEAT	279	312
FT	REPEAT	313	345
FT	REPEAT	346	378
FT	REPEAT	379	395
FT	DOMAIN	554	613
FT	DOMAIN	663	757
FT	DOMAIN	2104	2167
FT	DOMAIN	929	932
FT	DOMAIN	1010	1015
FT	DOMAIN	1022	1027
FT	DOMAIN	1194	1199
FT	DOMAIN	1850	1860
FT	VARSPIC	1	613
FT	VARSPIC	615	654
FT	VARSPIC	646	654
FT	VARSPIC	797	804
FT	VARSPIC	1930	1943
FT	VARSPIC	1944	2167
FT	CONFLICT	1141	1141
FT	CONFLICT	1174	1174
FT	CONFLICT	1246	1246
FT	CONFLICT	1323	1323
FT	CONFLICT	1331	1331
FT	CONFLICT	1726	1726
QO	SEQUENCE	2167 AA;	22633 MW; 3547B5A7B18BA6 CRC64;

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Comugen Ltd.

OM protein - protein search, using sw model

Run on: July 8, 2003, 10:58:52 ; Search time 80 seconds
(without alignments)
829,340 Million cell updates/sec

Title: US-10-018-706-2

Perfect score: 1670
Sequence: 1 MVTYTAINSQNKPIKRLGL.....LFEFRISRNGVYDPLTVLK 322

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_21:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp Vertebrate:*
14: sp Unclassified:*
15: sp_rvlinus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	268	15.0	314	2 Q9JP90	Q9JP90 rhodocyclus
2	253	15.1	242	16 Q9PP23	Q9PP23 xylella fas
3	249.5	14.9	467	16 Q9CCK1	Q9CCK1 pasteurella
4	238	14.3	259	16 Q8X2T2	Q8X2T2 escherichia
5	234	14.0	231	16 Q9HUP0	Q9HUP0 pseudomonas
6	232.5	13.9	250	16 Q8Z3Y0	Q8Z3Y0 salmonella
7	232.5	13.9	252	16 Q8ZM83	Q8ZM83 salmonella
8	228.5	13.7	269	2 Q9L528	Q9L528 pseudomonas
9	227.5	13.6	230	2 Q9K120	Q9K120 coxiella bu
10	226.5	13.6	244	2 Q9MNV6	Q9MNV6 pseudomonas
11	221	13.2	217	2 Q9JAG4	Q9JAG4 azobacter
12	220	13.2	268	16 Q8Y038	Q8Y038 ralistonia s
13	219.5	13.1	454	16 Q9RS36	Q9RS36 deinococcus
14	218.5	13.1	333	16 Q8ZBQ1	Q8ZBQ1 yersinia pe
15	208.5	12.5	311	16 Q9K1T7	Q9K1T7 vibrio chol
16	205	12.3	427	16 Q8YGS9	Q8YGS9 brucella me

17	203	12.2	415	16 Q9JTP1	Q9JTP1 neisseria m
18	202.5	12.1	379	16 Q8X7Z3	Q8X7Z3 escherichia
19	198	11.9	415	16 Q9JYVP	Q9JYVP neisseria m
20	196.5	11.8	515	16 Q9BUD1	Q9BUD1 rhizobium l
21	195	11.7	609	16 Q9A6T7	Q9A6T7 caulobacter
22	188.5	11.3	562	16 Q8UBQ5	Q8UBQ5 agrobacteri
23	174.5	10.4	512	2 Q08251	Q08251 rhizobium m
24	174	10.4	512	16 Q926D2	Q926D2 rhizobium m
25	172.5	10.3	392	2 Q8RNB5	Q8RNB5 bartonella
26	171.5	10.3	312	16 Q9X7W8	Q9X7W8 streptomyce
27	171	10.2	449	16 Q8RBQ3	Q8RBQ3 thermoaer
28	169	10.1	223	16 Q8XJL6	Q8XJL6 escherichia
29	167	10.0	371	16 Q83190	Q83190 treponema p
30	165.5	9.9	401	2 Q9KJW8	Q9KJW8 bartonella
31	158.5	9.5	436	16 Q8Y4E2	Q8Y4E2 listeria mo
32	158	9.5	321	16 Q8RG41	Q8RG41 fusobacteri
33	158	9.5	375	16 Q9RM21	Q9RM21 deinococcus
34	155	9.3	437	16 Q9Z7Y9	Q9Z7Y9 listeria in
35	153.5	9.2	176	2 Q52606	Q52606 ralistonia s
36	153	9.2	228	16 Q9X8M5	Q9X8M5 streptomyce
37	152.5	9.1	203	2 Q9X6S4	Q9X6S4 vibrio para
38	152	9.1	760	16 Q8YRU0	Q8YRU0 anabaena sp
39	147.5	8.8	176	2 Q9F7X9	Q9F7X9 escherichia
40	145	8.7	247	2 Q9S4T2	Q9S4T2 legionella
41	141	8.4	162	2 Q9R3D9	Q9R3D9 escherichia
42	141	8.4	164	2 Q9APL3	Q9APL3 escherichia
43	139.5	8.4	512	2 Q03491	Q03491 listeria gr
44	138.5	8.3	715	16 P74517	P74517 synchocyst
45	136.5	8.2	285	16 Q9PEP1	Q9PEP1 xylella fas

ALIGNMENTS

RESULT 1
Q9JP90 PRELIMINARY; PRT; 314 AA.
ID Q9JP90
AC Q9JP90;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE NLPD protein.
GN NLPD.
OS Rhodocyclus gelatinosus (Rhodopseudomonas gelatinosa).
OC Bacteria; Proteobacteria; beta subdivision; Comamonadaceae;
OC Rubrivivax.
OX NCBI_TaxID=28068;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=IL144;
RC MEDLINE=94132007; PubMed=8300574;
RA Nagashima K.V., Shimada K., Ohama S., Shimada K.;
RT "Phylogenetic analysis of photosynthetic genes of Rhodocyclus
gelatinosus: Possibility of horizontal gene transfer in purple
bacteria.";
RL Photosyn. Res. 36:185-191(1993).
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN=IL144;
RC MEDLINE=94132007; PubMed=8300574;
RA Nagashima K.V., Matsura K., Ohama S., Shimada K.;
RT "Primary structure and transcription of genes encoding B870 and
RT photosynthetic reaction center apoproteins from Rubrivivax
gelatinosus.";
RL J. Biol. Chem. 269:2477-2484(1994).
RN [3]
RP SEQUENCE FROM N.A.
RA STRAIN=IL144;
RC Igarashi N., Shimada K., Matsura K., Nagashima K.V.;
RT "Photosynthetic gene cluster in purple bacterium, Rubrivivax
gelatinosus." (eds.);
RL (in) Garab G. (eds.);
RT Photosynthesis:
mechanisms and effects (Proceedings of the 11th international congress

[1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=PW70;
 RC MEDLINE=21145866; PubMed=11248100;
 RX May B.J., Zhang Q., Li L.L., Paustian M.L., Whitam T.S., Kapur V.;
 RT "Complete genome sequence of *Pasteurella multocida* Fm70.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
 DR EMBL; AB006198; AAK03698.1; -
 DR InterPro; IPR002482; LysM.
 DR InterPro; IPR002886; Peptidase_M37.
 DR Pfam; PF01476; LysM; 2.
 DR SMART; SM00257; LysM; 2.
 DR Hypothetical protein; Complete proteome.
 KW HYPOTHETICAL 467 AA; 49735 MW; A98CDF70DD284A6 CRC64;
 SQ SEQUENCE 467 AA; 49735 MW; A98CDF70DD284A6 CRC64;
 Query Match 14.9%; Score 249.5; DB 16; Length 467;
 Best Local Similarity 30.0%; Pred. No. 3.6e-10;
 Matches 80; Conservative 42; Mismatches 110; Indels 35; Gaps 9;
 QY 73 YQYKQGVTSKIAQRYGLNREIGHINNNLSYTIYTGWLTLMSGDLKVEREISGSVN 132
 DB 215 YTRKQDPTMLIAYISGLDKVELASLNNMSEPRILSVQTLRVSGNG---RVASTSQPV 270
 QY 133 TAHTPSFVAVQSSRPVQOHPAVOKPTPP-----VVVYKPTPTPPVVOQAPAVA 182
 DB 271 T---QPVTVPVSGP--KSEBVTYTPGPHGQYGSDDTLIGPKSVSAAPVPVQPEPV 324
 QY 183 PPV--TEAPATGSSGVN---QFRYPVGAITNPVRFGRATVAGSTVTSNGMFGSRDG 236
 DB 325 KPVESTSVPVSTSSKIMVSNVTWQMPKGN--IVQGFSTADGG---NKGIDIASRG 377
 QY 237 DLINASAGTVIAQADHMDGAS--IVQHTNGFVSVYHIKDAQYKGTDTRTGQRTASGM 294
 DB 378 QAVNAAAGVAVVAGNLRGYNLITIKENDDYLSAVANESILVKKQGVRAQQQAKM 437
 QY 295 KNQPSGAALPEFRISRGVYVDPPLTVL 321
 DB 438 GSSGTSNVKLFELIRYKGSVDPTRYL 464
 RESULT 4
 Q8X2T2 PRELIMINARY; PRT; 259 AA.
 ID Q8X2T2 PRELIMINARY; PRT; 259 AA.
 AC Q8X2T2;
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Putative lipoprotein.
 GN BCS3738.
 OS *Escherichia coli* O157:H7.
 OC Bacteria; Proteobacteria; gamma subphylum; Enterobacteriaceae;
 OC *Escherichia*.
 OC NCBI_TaxID=83334;
 OX NCBI_TaxID=83334;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / RIMD 0509952;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayaishi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takami H., Honda T., Sasaki K., Ogasawara N., Yasunaga T.,
 RA Kohara S., Shiba T., Hattori M., Shinagawa H.;
 RT "Complete genome sequence of enterohemorrhagic *Escherichia coli*
 RT O157:H7 and genomic comparison with a laboratory strain K-12.";
 RT DNA Res. 8:11-22(2001).
 RL EMBL; AB002563; BAB37161.1; -
 DR InterPro; IPR002482; LysM.
 DR InterPro; IPR002886; Peptidase_M37.
 DR Pfam; PF01476; LysM; 1.
 DR Pfam; PF01551; Peptidase_M37; 1.
 DR SMART; SM00257; LysM; 1.
 KW Lipoprotein.
 SQ SEQUENCE 259 AA; 27528 MW; 8A1B89E01F9B72E8 CRC64;

Query Match 14.3%; Score 238; DB 16; Length 259;
 Best Local Similarity 24.9%; Pred. No. 1.1e-09;
 Matches 78; Conservative 46; Mismatches 103; Indels 86; Gaps 8;
 QY 16 KRGLRFGVITTCILACGASKPPTNYSGGSHRTSSGGGLAGISQVITTSQGVNRYQV 75
 DB 17 KSLIGVWLLSVGILLGAC-----SGSKSPDTGYSG-----SVYTV 52
 QY 76 KQDPTVSKIAQRYGLNREIGHINNNLSYTIYTGWLTLMSGDLK---VRESISGSV 131
 DB 53 KRQDPTLRISRTTGTSTVKEALRLNGISPPYITIEGQTLK--GSAKSSSTRKSTAKSTT 110
 QY 132 NTAH-TPSPVAVQSSRPVQOHPAVOKPTPPVVKKPTPTPPVVOQAPAVAPVTEAPF 190
 DB 111 KTASVTPSSAVPSSSWPVPVQRCMLMPTTGKVLIM-----PY 146
 QY 191 ATGSSGVMOFRYPVGAITNPVRFGRATVAGSTVTSNGMFGSRDGLINASNAVTYQA 250
 DB 147 STADGG-----NKGIDIASRGTPIYAAGAKVYV 177
 QY 251 DHNMDGAS--IVQHTNGFVSVYHIKDAQYKGTDTRTGQRTASGMKQPSGALFEPR 308
 DB 178 GNQLRGYGNLIMTKHSEDTYTAHANDTMLVNNQGSYKQAKLATWGSTDAASVRLHFOI 237
 QY 309 SRNGVYVDPPLTVL 321
 DB 238 RYRATADPLRYL 250
 RESULT 5
 Q9HUP0 PRELIMINARY; PRT; 231 AA.
 ID Q9HUP0 PRELIMINARY; PRT; 231 AA.
 AC Q9HUP0;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Hypothetical protein PA4924.
 GN PA4924.
 OS *Pseudomonas aeruginosa*.
 OC Bacteria; Proteobacteria; gamma subphylum; Pseudomonadaceae;
 OC *Pseudomonas*.
 OC NCBI_TaxID=287;
 OX NCBI_TaxID=287;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 15692 / PAOI;
 RX MEDLINE=20437337; PubMed=10984043;
 RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
 RA Hickey M.J., Brinkman F.S.L., Huftagle W.O., Kowalik D.J., Lagrou M.,
 RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
 RA Reizer U., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
 RT "Complete genome sequence of *Pseudomonas aeruginosa* PAOI, an
 RT opportunistic pathogen.";
 RT Nature 406:959-964(2000).
 RL Nature 406:959-964(2000).
 DR EMBL; AB004905; AAG08309.1; -
 DR HSP; P23931; IEOG.
 DR InterPro; IPR002482; LysM.
 DR InterPro; IPR002886; Peptidase_M37.
 DR Pfam; PF01476; LysM; 1.
 DR Pfam; PF01551; Peptidase_M37; 1.
 DR SMART; SM00257; LysM; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 231 AA; 25174 MW; 3FC8EAB82A42B91 CRC64;
 Query Match 14.0%; Score 234; DB 16; Length 231;
 Best Local Similarity 26.7%; Pred. No. 1.9e-09;
 Matches 68; Conservative 37; Mismatches 90; Indels 60; Gaps 6;
 QY 69 VPNRYQKQDPTVSKIAQRYGLNREIGHINNNLSYTIYTGWLTLMSGDLKVEREIS 128
 DB 32 VRGRYVYKQDPTVSKIAQRYGLNREIGHINNNLSYTIYTGWLTLMSGDLKVEREIS 90

QY		129	SGVNTAHTPSSEVAAGSSRRPVOQHADVCKETPPVVVVKKCTPTTPPVQQCPAPAVAPVTBA	188
Dd		91	SSSNTR-----ARKPPPPPEVT-----	108
QY		189	PFAATGSGIMQFRYPVGATNPVVRFGTATVAGSTVTSNGMFMFSGDGDILINASNACTVI	248
Dd		109	-----LRQMGMFM--KGPVIRRF-----SSSDRLNKIGRIHAGTLGGPQVASLAKGVV	153
QY		249	QADHNMDGAS--IIVQHTNGPFVSSYLIIHKDAQVKJGDTVRTGORISMKNQSPGALLFEF	306
Dd		154	FAYNNMRKYGNLVITOHGHSYSTSTAHNRRLLVKGSMYGKQKLTAAGSSDDADRQVLVF	213
QY		307	RISRNGVVVDPLTVL	321
Dd		214	EIRQNGRFLDPUSLL	228

RESULT 6

ID	Q823Y0;	PRELIMINARY;	PRT;	250 AA.
AC	0823Y0;			
DT	01-MAR-2002 (TrEMBLrel. 20, Created)			
DT	01-MAR-2002 (TrEMBLrel. 20, last sequence update)			
DT	01-JUN-2002 (TrEMBLrel. 21, last annotation update)			
DE	Possible lipoprotein.			
GN	STR3194.			
OS	Salmonella typhi.			
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;			
OC	Salmonella			
OX	NCBI_Taxid=601;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CT18;			
RX	MEDLINE=15134947; PubMed=11677608;			
RA	Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,			
RA	Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahina M.,			
RA	Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,			
RA	Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,			
RA	Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,			
RA	Krogsh A., Larsen T.S., Leather S., Moulie S., O'Garra P., Parry C.,			
RA	Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,			
RA	Whitehead S., Barrrell B.G.;			
RT	"Complete genome sequence of a multiple drug resistant Salmonella			
RT	enterica serovar Typhi CT18."			
RL	Nature 413:848-852(2001).			
DR	EMBL: AL627277; CAD02868.1; --			
DR	InterPro: IPR002482; LYSM.			
DR	InterPro: IPR002886; Peptidase_M37.			
DR	Pfam: PF01476; LYSM.1.			
DR	Pfam: PF01551; Peptidase_M37.1.			
DR	SMART: SM00257; LYSM.1.			
DR	Lipoprotein; Complete proteome.			
QO	SEQUENCE 250 AA; 26426 MW; 9A57CF6117BE76C4 CRC64;			

Query Match	13.9%	Score 232.5	DB 16	Length 250
Best Local Similarity	24.4%	Pred No. 2.7e-09		
Matches	76	Conservative	49	Mismatches 99
			Indels	87
			Gaps	9
QY	18	LGILFGVITTCIAGCASKPTYNSTSGSGSHRTSSGGLAISQVYITSGGVNFRQYQK	77	
Db	11	LGIAVLCTGELLAC-----SSNSGGQYV---SGS-----VYTYKR	44	
QY	78	GDTVSKIAQRYGLMNRKEIGHININLSSYTYITGQWLTMISGDLK---VREKSISSGVNT	133	
Db	45	GDTLYIRISRACTSVKEIARLNLGISPPYITVEVQGIKY-RGSAKSSSTRKTSNKTAKTI	103	
QY	134	AHT--PSPVAVOSSRPVNOQHPAVQCKPTPPVYVKKPCTPTPPVQQPAVAPRVTEAPAT	192	
Db	104	AAVRPSSSVFSSWPPVQRCWYMWANGVIL-----PST	139	
QY	193	GSSGVMOFRYPVGAATNPVRPFGTATVAGSTVSNQMWFSGEDGLINASNAGTVIQADH	252	

[illegible]

RESULT 7

ID	PRELIMINARY;	PRT;	252 AA.
AC	Q8ZM63;		
DT	01-MAR-2002 (TrEMBLrel. 20, Created)		
DT	01-MAR-2002 (TrEMBLrel. 20, Last sequence update)		
DT	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)		
DE	Putative metalloendopeptidase.		
OS	STM3038.		
OS	Salmonella typhimurium.		
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;		
OC	Salmonella.		
OX	NCBI_TaxID=602;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	SMRAIN-LT2 / SGSC412 / ATCC 700720;		
RC	MEDLINE=21534948; PubMed=11677609;		
RA	McClintland M., Sanderson K.E., Spiegh J., Clifton S.W., Latreille P.		
RA	Courtney L., Porciliak S., Ali J., Dante M., Du F., Hou S., Layman D.		
RA	Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,		
RA	Ryan E., Sun H., Flores L., Miller W., Stonering T., Nhan M.,		
RA	Waterston R., Wilson R.K.;		
RI	Complete genome sequence of <i>Salmonella enterica</i> serovar Typhimurium		
RI	LT2.		
RL	Nature 413:852-856(2001).		
DR	EMBL; AEO08835; AAL21913.1; --		
DR	InterPro; IPR002482; LysM.		
DR	InterPro; IPR002886; Peptidase_M37.		
DR	Pfam; PF01476; LysM; 1.		
DR	Pfam; PF01551; Peptidase_M37; 1.		
DR	SMART; SM00257; LysM; 1.		
DR	Hypothetical protein; Complete proteome.		
QO	SEQUENCE 252 AA; 26652 MW; 6ADB04D76A8E61E CRC64;		

Query Match	13.9%	Score 232.5	DB 16	Length 255
Best Local Similarity	24.4%	Pred. No. 2.8e-09		
Matches	76	Conservative 49	Mismatches 99	Indels 87
			Gaps	9
QY	18	LGILFGVTTTCILGACASKPFTYNTSGSGSRTSGSGGLAIGSYITDSQGVNREYQVQ	77	
DB	13	LGIAVLITCTGLLAGC-----SSNSGSGTY--SGS-----VYTVKR	46	
QY	78	GDTYSKIAQORGLMMREIGHINININSSTYITGOWLTMSGDLK---VREIRISSGVNT	133	
DB	47	GDTLYIRISRACTSVKELARLNGISPTTIEVGQRIKX-RSGAKSSTSTRKTSNKTATXT	105	
QY	134	AHT--PSPAVQSSPYPVOOHFAVQKPPPPVVVVKKPTPTPPVQGPAPVAPVPTBAPPT	192	
DB	106	AAVRPSSSVSPSSMPVQRCWMPVNPANKVLL-----PYST	141	
QY	193	GSSGVMQRYVVGATNPVYRRFGATVAGSYTNSGMWFSGGEDDLINASNGYIQADH	252	
DB	142	AEGG-----NKGDIDLAGRGFPVVAAGAKGVYVEN	172	
QY	253	NMDGAS--IYVQHTNGFVFSYTHIKDAQVKTGDPYRTQRTQASMNQSGALLFFERIS	310	
DB	173	QLRGKGNILMIKHNEDYITTAIYAHNDYMLVNNQGSYVDAQKTAIAGSTDAASVRLHFQIKX	232	
QY	311	NGVYVDPPLTVL	321	
DB	233	RATADIDPLRYL	243	


```

RESULT 8
ID 09L528 PRELIMINARY; PRT; 269 AA.
AC 09L528;
DT 01-OCT-2000 (TEMBLrel. 15, Created)
DT 01-OCT-2000 (TEMBLrel. 15, last sequence update)
DE 01-JUN-2001 (TEMBLrel. 17, last annotation update)
GN NLPD.
OS Pseudomonas putida.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=303;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KT2440;
RA Ojangu E., Tover A., Kivisaar M.;
RT "Sequence of Pseudomonas putida nlpd gene (complete sequence) and ORF
RT similar to E. coli pcm gene (partial sequence).";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF260132; AAF70311.1; -.
DR InterPro: IPR002482; LysM.
DR InterPro: IPR002886; Peptidase_M37.
DR Pfam: PF01476; LysM; 1.
DR Pfam: PF01551; Peptidase_M37; 1.
DR SMART: SM00257; LysM; 1.
SQ SEQUENCE 269 AA; 2836 MW; F3F5DCAB481B6B14 CRC64;

Query Match 13.7%; Score 228.5; DB 2; Length 269;
Best Local Similarity 29.2%; Pred. No. 5.6e-09;
Matches 81; Conservative 36; Mismatches 95; Indels 65; Gaps 12;

QY 29 ILAGCASKPTYNSTSGSGSRISGGLAIGSQYITTSQGVNPR-----YQKQSDTV 81
DB 4 LTVGCS-----TSN-----SARVDDNNNTVPKPAVTSGQYIVKPDTL 44
QY 82 SKIAQRYGLMREIGHINNINSYTIYTGWLTLMSGDLKVERSISSGVNTAHTPPVA 141
DB 45 FSIAPRYGMXYKELAAANGIOAPYITPQAIRSSSS-----SRRTTVSSPS--- 93
QY 142 VQSSRPVQHPAVQKPTPPVVVVKPTPTPPVQVQAPVAPVTEAPFATGSSGVNQFR 201
DB 94 -SSRTVTTRPVGSTATATASTSKPATSA-----SAPAPVAVTVAERAVG-----WT 144
QY 202 YPVGATNPVRRRGTAIVASTVTSNGMSPSGRD--GDL--INASAGYVIOADHMDG 256
DB 145 WP--ANGVLLGK-----ASNGSLNKGIDIDGLQPVPAADGAVVYVAGSGLRG 192
QY 257 AS--IVIOHTNGFVSYIHIKDAQVKTGDTVTRTGQRI 291
DB 193 YGELIITIKHSDTVSAVGHNRRLVREGQVYKAGQSI 229

RESULT 9
ID 09K120 PRELIMINARY; PRT; 230 AA.
AC 09K120;
DT 01-OCT-2000 (TEMBLrel. 15, Created)
DT 01-OCT-2000 (TEMBLrel. 15, last sequence update)
DE 01-DEC-2001 (TEMBLrel. 19, last annotation update)
DE Lipoprotein precursor NLPD.
OS Coxiella burnetii.
OC Bacteria; Proteobacteria; gamma subdivision; Legionellaceae group;
OC Coxiella group; Coxiella.
OX NCBI_TaxID=777;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21340372; PubMed=11447163;
RA Seshadri R., Samuel J.E.;
RT "Characterization of a Stress-Induced Alternate Sigma Factor, Rpos, of
RT Coxiella burnetii and Its Expression during the Development Cycle.";
RL Infect. Immun. 69:4874-4883 (2001).

```

```

DR EMBL: AF244357; AAF73517.1; -.
DR InterPro: IPR002482; LysM.
DR InterPro: IPR002886; Peptidase_M37.
DR Pfam: PF01476; LysM; 1.
DR Pfam: PF01551; Peptidase_M37; 1.
DR SMART: SM00257; LysM; 1.
SQ SEQUENCE 230 AA; 25296 MW; 48312658B04F8533 CRC64;

Query Match 13.6%; Score 227.5; DB 2; Length 230;
Best Local Similarity 28.5%; Pred. No. 5.6e-09;
Matches 72; Conservative 31; Mismatches 83; Indels 67; Gaps 8;

QY 73 YQVKGQDPTVSKIAQRYGLMREIGHINNINSYTIYTGWLTLMSGDLKVERSISSGVN 132
DB 40 YRVKQGDPTVSIAMAFGLDPRALAAANRLSPRIEAGQTLQW-----TIPRG-- 88
QY 133 TAHTPSFVAVQSSRPVQHPAVQKPTPPVVVVKPTPTPPVQVQAPVAPVTEAPFAT 192
DB 89 -AHVSGRFA-----AAPRQSPPLQP----- 109
QY 193 GSSGVNQFRYPVGTATNPVRRFGTATV--AGSTVTSNGMSPSGRDGDLINASNAGTVIOA 250
DB 110 ----VAHRWPANG-----RLIGYSVMAG-----NIGINIAGYGAIVRAASGVVYS 156
QY 251 DHNMDGAS--IVIOHTNGFVSYIHIKDAQVKTGDTVTRTGQRIASMKNOPSGALFEFRI 308
DB 157 GAGIRGYGNLIYKHHTTYSIAVAFNKRVLVKGSSRVRAQKIAEMGRTSGRYMLHFEI 216
QY 309 SRNGVYVDPLTVL 321
DB 217 RRGQPVNPLRYL 229

```

```

RESULT 10
ID 09MWV6 PRELIMINARY; PRT; 244 AA.
AC 09MWV6;
DT 01-NOV-1999 (TEMBLrel. 12, Created)
DT 01-NOV-1999 (TEMBLrel. 12, last sequence update)
DE 01-DEC-2001 (TEMBLrel. 19, last annotation update)
DE Lipoprotein (Fragment).
GN NLPD.
OS Pseudomonas putida.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=303;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WCS358;
RX MEDLINE=20135599; PubMed=10673044;
RA Kotlic M., Degraess G., Venturi V.;
RT "Cloning and characterisation of the rpos gene from plant growth-
RT promoting Pseudomonas putida WCS358: Rpos is not involved in
RT siderophore and homoserine lactone production.";
RL Biochim. Biophys. Acta 1489:413-420 (1999).
DR EMBL: Y19122; CAB46190.1; -.
DR InterPro: IPR002482; LysM.
DR InterPro: IPR002886; Peptidase_M37.
DR Pfam: PF01476; LysM; 1.
DR Pfam: PF01551; Peptidase_M37; 1.
DR SMART: SM00257; LysM; 1.
FT NON TER 1
SQ SEQUENCE 244 AA; 25191 MW; 6F9B5BA39CD739AE CRC64;

Query Match 13.6%; Score 226.5; DB 2; Length 244;
Best Local Similarity 28.5%; Pred. No. 7.2e-09;
Matches 82; Conservative 43; Mismatches 96; Indels 67; Gaps 13;

QY 20 LIRGVITTCILACGAKPTYNSTSGSGSRISGGLAIGSQYITTSQGVNPR----- 72
DB 3 LVIALAGTTLACGSSR--SSTS-----ARVVDNRNNAAPKPPYVTSQ 43
QY 73 YQVKGQDPTVSKIAQRYGLMREIGHINNINSYTIYTGWLTLMSGDLKVERSISSGVN 132

```



```

Complete proteome.
SQ SEQUENCE 311 AA; 33507 MW; FC70D9EF1396451E CRC64;

Query Match 12.5%; Score 208.5; DB 16; Length 311;
Best Local Similarity 25.8%; Pred. No. 2e-07;
Matches 85; Conservative 48; Mismatches 133; Indels 59; Gaps 12;

QY 17 RLGLIFGITTICILAGCASKETYNSTSGSGSHRTSGSGLAIGSVITDSQGVNRYQVK 76
DB 7 RLGLL--LFCSLFEGCTA-PIPAVPSGLKQDYNKVEKSGYRGSF-----YEVK 51
QY 77 QGDTVSKIAQRYGRLNMRFEIGHINNINSYITYGWMLTLM-----SGDLKY-- 122
DB 52 KGDILYFAIYLTDKQVNDLISYNDLAPPTIHPGKIKLWLPNTPPAYGCTGATVAV 111
QY 123 --RERSISSGVNTAHPSPVAVOGSSRPVQ-----QHPAVQKPTPPVVKKPTPRP 173
DB 112 ASSTASAVAKAATATVATQVTSKSNNTKVNQNGSTNSQNLTKQDVKTDQTKREYEP 171
QY 174 VVOQPAFY---APVTEAPFATGSSGVMOFRYPVGATNPVRRFGTATVAGSTVTSNGM 229
DB 172 VQKONVNVNVAKAKPSDEK-----IAKWLMP--TKGRVYKNFS---AGDQ-GNKGI 216
QY 230 WFSGRDGLINASNAGTVIOADHNM DGAS--IVIOHTNGFVSSYTHIKDAQVKTGDTVRT 287
DB 217 DIAGQGGAVVAATDAGTVVYSGNALRGYGNLIHIGNEHYLSAVAHNDQLAKEGQTVQA 276
QY 288 GORIASMKNQPSGALFEFRISRNQVYVP 317
DB 277 GOKIATMGSSGTNSVRLHFEIRYQCKSYNP 306

RESULT 16
08YGS9 PRELIMINARY; FRT; 427 AA.
AC Q8YGS9;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Lipoprotein NLPD.
GN BMEI1079.
OS Brucella melitensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Brucellaceae; Brucella.
OX NCBI_TaxID=29459;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=16M / ATCC 23456 / BIOTYPE 1;
RX MEDLINE=20020109; PubMed=11756688;
RA DeLvecchio V.G., Kaprethel V., Redkar R.J., Patra G., Mujar C., Los T.,
RA Ivanova N., Andersen I., Bhattacharyya A., Lykidis A., Resnik G.,
RA Jalonoski L., Larsen N., D'Souza M., Bernal A., Mazur M., Goldstein E.,
RA Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Leetson J.-J.,
RA Hasekorn R., Kyriades N., Overbeck R.;
RT "The genome sequence of the facultative intracellular pathogen
RT Brucella melitensis."
RL Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).
DR EMBL; AE009547; ALN52260.1; -.
DR InterPro; IPR002482; LysM.
DR InterPro; IPR002886; Peptidase_M37.
DR Pfam; PF01476; LysM; 1.
DR Pfam; PF01551; Peptidase_M37; 1.
DR SMART; SM00257; LysM; 1.
KW Complete proteome.
SQ SEQUENCE 427 AA; 43604 MW; 48912D48846EBD7 CRC64;

Query Match 12.3%; Score 205; DB 16; Length 427;
Best Local Similarity 26.2%; Pred. No. 5.4e-07;
Matches 79; Conservative 45; Mismatches 132; Indels 46; Gaps 12;

QY 30 LAGCASKETYNSTSGSGSHRTSGSGLAIGSVITDSQGVNRYQVKQDGYTSKIAQRYG 89
DB 162 IAGVQAPAVNAGKSSPTNMAAGCAIA-----TPSVGAVVYVKGSDLSIFSIAKKN 214

```

```

QY 90 LMNREIGHINNINSYITYGOMLTLSGDLKVEREISSGVNT---AHTSPV-AVOSS 145
DB 215 VPVQDLKRAAGLTSG-AIRVQOSLVI-----PSAAGAPFOVAITFPQANPKAA 264
QY 146 RPPVQAPVQKPPPPVNVVKKPTPTPPVQAPAPVPTAPFATGSSGVMOFRYPVG 205
DB 265 SAPANGAFAVYKPTFPQASNK-----VIEDAE-----KQQAAPSSTISQRMFPY- 310
QY 206 ATNPVRRFGATVAGSTVTSNGMFGSGRDDLNASNAGVIOADHNM--GASIVQH 263
DB 311 -RGLMSFGQRE--GTSVSDGIDIVPEETPYKALENGVLITAGGGLKEFGQTLIRH 366
QY 264 TNGFVSSYTHIKDAQVKTGDTVRTGQRIASMKNQPSGAA--LFEFRISRNQVYVPLTV 320
DB 367 DNGLVTVYGHNSQITVQVQKVRKGEELA--KSGMSGNAKSPKLHFEVRKNSAFVNPYK 424
QY 321 LK 322
DB 425 LE 426

RESULT 17
09UTP1 PRELIMINARY; FRT; 415 AA.
AC 09UTP1;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Putative membrane peptidase.
GN NMA1692.
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=65699;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Z2491 / SEROGROUP A / SEROTYPE 4A;
RX MEDLINE=20222556; PubMed=10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Davies R.M., Davis P., Devlin K., Feltham T., Hamlin N., Holroyd S.,
RA Jagers K., Leather S., Moule S., Mungall K., Quail M.A.,
RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA Whitehead S., Spratt B.G., Barrall B.G.;
RT "Complete DNA sequence of a serogroup A strain of Neisseria
RT meningitidis Z2491."
RL Nature 404:502-506(2000).
DR EMBL; AL162756; CAB84920.1; -.
DR InterPro; IPR002482; LysM.
DR InterPro; IPR002886; Peptidase_M37.
DR Pfam; PF01476; LysM; 2.
DR Pfam; PF01551; Peptidase_M37; 1.
DR SMART; SM00257; LysM; 2.
KW Complete proteome.
SQ SEQUENCE 415 AA; 42883 MW; B39BBA52C7CD78F7 CRC64;

Query Match 12.3%; Score 203; DB 16; Length 415;
Best Local Similarity 27.3%; Pred. No. 7.2e-07;
Matches 71; Conservative 42; Mismatches 97; Indels 50; Gaps 11;

QY 73 YQVQKQDTVSKIAQRYGRLNMRFEIGHINNINSYITYGOMLTLSGDLKVEREISSGVN 132
DB 185 HTVIRGDTVNVISKRHYISODDFPAMGMNTN-RLTSGQ-----IVKVPAGYAA-- 233
QY 133 TAHTSPVAVOGSSRPVQOHAVQKPTPPVNVVKKPTPTPPVVOQAPVAD----- 183
DB 234 ----PKAAAVK-SRPAPV--AAVQTPV-----KPAQPPVQSAPOFAAANAKVAPA 279
QY 184 PUTEAPFA---TSSSGVMQFRYPVGATNPVRRFGTATVAGSTVTSNGMWFSGRDGLI 239
DB 280 PAPQSPAPSPGSTRSGVGIWQRP--TQGKVVAAPFGG-----NNKGVDIAGNAGQPV 329
QY 240 NASNAGTVIOADHNM DGAS--IVIOHTNGFVSSYTHIKDAQVKTGDTVRTGQRIASMKNQ 297

```

Db 330 LAADGCVYVAGSGLRGYGNLVITIOHNSFLTAIGHNOKLVBEGQGVKGGQVALMGNT 389
 Qy 298 PSGALFEFRISRNQVYDP 317
 Db 390 EASRTQLHFEVRONGKFPVNP 409

RESULT 18

Q98723 PRELIMINARY; PRT; 379 AA.
 ID Q98723; PRELIMINARY; PRT; 379 AA.
 AC Q98723; PRELIMINARY; PRT; 379 AA.
 DT 01-MAR-2002 (Tremblrel. 20, Created)
 DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Lipoprotein.
 GN NLPD OR Z4050 OR ECS3596.
 OS Escherichia coli O157:H7.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxId=83334;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 Postfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 Grobeck J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
 Apodaca E., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 Welch R.A., Blattner F.R.;
 RA "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";
 RT Nature 409:529-533(2001).
 RL [2]
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / RIMD 0509952;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayashi T., Makino K., Onishi M., Kurokawa K., Ishii K., Yokoyama K.,
 Han C.-g., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 Iida T., Takami H., Honda T., Sasaki K., Ogasawara N., Yasunaga T.,
 Kuhara S., Shiba T., Hattori M., Shinagawa H.;
 RA "Complete genome sequence of enterohaemorrhagic Escherichia coli
 RT O157:H7 and genomic comparison with a laboratory strain K-12";
 RL DNA Res. 8:11-22(2001).
 DR EMBL; AE005502; AAG57849.1; -
 DR EMBL; AP002562; BAB37019.1; -
 DR InterPro; IPR002482; LysM.
 DR InterPro; IPR002886; Peptidase_M37.
 DR Pfam; PF01476; LysM; 1.
 DR Pfam; PF01551; Peptidase_M37; 1.
 DR SMART; SM00257; LysM; 1.
 KW Complete proteome.
 SQ SEQUENCE 379 AA; 40172 MW; F664459AID90720 CRC64;

Query Match 12.1%; Score 202.5; DB 16; Length 379;
 Best Local Similarity 26.1%; Pred. No. 7e-07;
 Matches 72; Conservative 42; Mismatches 113; Indels 49; Gaps 9;

Qy 73 YQVKGQDVTSKIAQRYGLNMRREIGHNINLNSSTIYTGQWLTMSGDLKYRERSISSGVN 132
 Db 123 YTVKQDITLFIYIMTIGNDFRDLAQRNINQAPYALNVGQ--TLQVG-----N 167
 Qy 133 TAHTPSVAVQSRPPVQOHBAVQKP--TPVVVVVKKPT-----P 170
 Db 168 ASGTPTTGGNATIQADAAEGGVVVKPQNSTVAVASQPTTYSSESGEGANMLPNK 227
 Qy 171 TTPVVOQAPVAPVTEAPFATG---SSGVMOFRYPGATNPVVRFGATVAGSTVTSN 227
 Db 228 TATTVTAAPVTPVPASTBPTVSSTSTPSTWRP--TEGKVIETFG-ASEGG---NK 280
 Qy 228 GMMFSGRQDGLINASNAQVITQADHNMDS--IVIQHTNGFVSSYTHIKDAQVKTGDTV 285
 Db 281 GIDIASGSKGQAITATADGRVYVAGNALRGYGNLIIIKHNDYLSAVAHNDTMLVREQEV 340

Qy 286 RTGQRASMKNOPSGALFEFRISRNQVYDP 321
 Db 341 KAGOKIATMGSTGTSITRLHFEIRYKKSIVNPLRYL 376

RESULT 19

Q94YP9 PRELIMINARY; PRT; 415 AA.
 ID Q94YP9; PRELIMINARY; PRT; 415 AA.
 AC Q94YP9; PRELIMINARY; PRT; 415 AA.
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Lipoprotein NlpD, putative.
 GN NMB1483.
 OS Neisseria meningitidis (serogroup B).
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
 OX NCBI_TaxId=491;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MC58 / SROGROUP B.
 RX MEDLINE=20175755; PubMed=10710307;
 RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
 Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
 Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
 Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
 Mason T., Ciecho J., Parksey D.S., Blair E., Citron E.H., Clark E.B.,
 Cotton M.D., Uetreback T.R., Khouiri H., Qin H., Yamachyan J.,
 Gail J., Scarlato V., Maserani V., Plaza M., Grandi G., Sun L.,
 Smith H.O., Fraser C.M., Moxon E.R., Rappoli R., Venter J.C.;
 RA "Complete genome sequence of Neisseria meningitidis serogroup B strain
 RT MC58";
 RL Science 287:1809-1815(2000).
 DR EMBL; AE002497; AAP41839.1; -
 DR TIGR; NMB1483; -
 DR InterPro; IPR002482; LysM.
 DR InterPro; IPR002886; Peptidase_M37.
 DR Pfam; PF01476; LysM; 2.
 DR Pfam; PF01551; Peptidase_M37; 1.
 DR SMART; SM00257; LysM; 2.
 KW Complete proteome.
 SQ SEQUENCE 415 AA; 42875 MW; 7704BF7F0D54DP5 CRC64;

Query Match 11.9%; Score 198; DB 16; Length 415;
 Best Local Similarity 27.2%; Pred. No. 1.7e-06;
 Matches 72; Conservative 38; Mismatches 95; Indels 60; Gaps 12;

Qy 73 YQVKGQDVTSKIAQRYGLNMRREIGHNINLNSSTIYTGQWLTMSGDLKYRERSISSG-- 130
 Db 185 HTIVRGDVTYNIISKRY-----HISQDD-----FRAMNG---MTDNLISIGOI 223
 Qy 131 --VNTAHTPSP--VAVOSSRPVQOHBAVQKPPTPPVVVKKPTPTPPVQOAPAPVAP--- 183
 Db 224 VKVPAPOSAPKPAVAESRPAPV--AAVQTPV-----KPAQPPVQSAQPPAPAPAE 274
 Qy 184 -----VTTAPFA-----TGSSGVMOFRYPGATNPVVRFGATVAGSTVTSNGMMSGR 234
 Db 275 KAVPAPAPQSPAPSPGTSVSGIIVWRP--TGCKVADBGG-----NNKGVDIAGN 324
 Qy 235 DGDGLINASNAQVITQADHNMDS--IVIQHTNGFVSSYTHIKDAQVKTGDTVTRGQRIA 292
 Db 325 AGQPVLAADGKVVYVAGSGLRGYGNLVITIOHNSFLTAIGHNOKLVBEGQVKKGGQVA 364
 Qy 293 SMKNQPSGALFEFRISRNQVYDP 317
 Db 385 LMGNTDASRTQLHFEVRONGKFPVNP 409

RESULT 20

Q98LD1 PRELIMINARY; PRT; 515 AA.
 ID Q98LD1; PRELIMINARY; PRT; 515 AA.
 AC Q98LD1; PRELIMINARY; PRT; 515 AA.
 DT 01-OCT-2001 (Tremblrel. 18, Created)

RP SEQUENCE FROM N.A.
 RX MEDLINE=21608551; PubMed=11743194; Miller N., Blanchard M.,
 RA Goodner B., Hinkle G., Gattung S., Cao Y., Akenazi M., Halling C., Mullin L.,
 RA Ourullo B., Goldman B.S., Gordon J., Vaudin M., Iartchouk O., Dep A., Liu P.,
 RA Houtmel K., Allinger J., Crowell C., Garsen J., Lomo C., Sear C., Strub G.,
 RA Flanagan C., Slaters S.,
 RA Cielo C., Slaters S.,
 RT "genome sequence of the plant pathogen and biotechnology agent
 RT Agrobacterium tumefaciens C58.";
 RL Science 294:2323-2328(2001).
 DR EMBL; AE009126; AA42700.1; ALT_INIT.
 DR EMBL; AE008091; AK87473.1; -.
 KW Complete proteome.
 SQ SEQUENCE 562 AA; 57218 MW; B6479E31F9579A84 CRC64;

Query Match 11.3%; Score 188.5; DB 16; Length 562;
 Best Local Similarity 25.1%; Pred. No. 1.2e-05;
 Matches 81; Conservative 64; Mismatches 131; Indels 47; Gaps 15;

QY 5 IAINSONOKPIKRLGLFGVITTCILAGCAKPTYNSTSGSHRTSGGGLA-IGSQVI 63
 DB 272 IALNRNDGP-----TPLRAPDGNVAVLPQAAARDKLSEAGKLTPEGKPL 319
 QY 64 TDSQGVNRYQVKQGDVTSKIAQRYGLMREIGHINNINSYTYTGQWLTLMGDLKVR 123
 DB 320 PPSGG-----YKVGQDGLAKIARANGVSVALKAANGI-SNESIRVGOTLTM----- 366
 QY 124 ERSSISGVNTAHTSPVAVQSSRPVQOHFVQKPTP--PVVVKKPTPPPVQOPAP 180
 DB 367 PGASTDITKTPASVAKKA--AAKPEV--TASIKPEPKAPAAATTAAPTPTASVSD 422
 QY 181 VAPVTEAPFATGSSGVNQFERYV-GATNPVRRFGTATVAGSVTSGMFGSRDGLI 239
 DB 423 IEKKADNASTAPSTIGIKRMPVPGA--VINNFGD-NVSGS--KNDGINISVDEGPI 476
 QY 240 NASNAGTVIOADHNM--GASIVIOHTNGFVSYIHIDKADQVKTGTVRTQRIASMKQ 297
 DB 477 KAANGVVIVYAGNGLKQLGNTVLVRHDDGKVTGVGNANLNVQGGVQVROQTAT--SG 534
 QY 298 PSGAA---LFEFRISRNGVYVDP 317
 DB 535 MTGSAKRPQVHFVVRKDATPVNP 557

RESULT 23
 008251 PRELIMINARY; PRT; 512 AA.
 ID 008251
 AC 008251;
 DT 01-JUL-1997 (Tremblrel. 04, Created)
 DT 01-JUL-1997 (Tremblrel. 04, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Lipoprotein.
 GN LPPB.
 OS Rhizobium meliloti (Sinorhizobium meliloti).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Sinorhizobium.
 OC NCBI_TaxID=382;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1021;
 RX MEDLINE=97449848; PubMed=9304864;
 RA Streit W.R., Phillips D.A., in a possible survival operon of
 RT Rhizobium meliloti.";
 RL Mol. Plant Microbe Interact. 10:933-937(1997).
 DR EMBL; U81296; AAB88077.1; -.
 DR InterPro; IPR002482; LysM.
 DR InterPro; IPR002886; Peptidase_M37.
 DR Pfam; PF01476; LysM; 2.
 DR Pfam; PF01551; Peptidase_M37; 1.
 DR SMART; SM00257; LysM; 1.
 SQ SEQUENCE 512 AA; 52431 MW; 0A870A76BEC7D8B CRC64;

Query Match 10.4%; Score 174.5; DB 2; Length 512;
 Best Local Similarity 23.5%; Pred. No. 0.00011;
 Matches 95; Conservative 58; Mismatches 128; Indels 123; Gaps 19;

QY 24 VITTCILAGCAKPTYNSTSGSHRTSGGGLAIGSQVITDSQGVNRYQVKQGDVTSK 83
 DB 127 VQTASLSQAAALPKAPPLAGAKD-MSGKGWS-----ASNAPT-INVROGDTVTY 176
 QY 84 IACRYGLMREIGHINNINSYTYTGQWLT-----SGVNTAHTPSP----- 113
 DB 177 LARFVGEKEILKANGLKASQVEPGQRLVPTFTAGSAKAASGS1ADVEGKKRP 236
 QY 114 -----TLMGDLKVERSSIS----- 139
 DB 237 SPLPTDQREVAAILPGQSQSRKESRSDVAAGKINSAGEGNGNAYTVKPDLSNR1AKA 296
 QY 140 -----VAQSSRP--PVQOHF--AVQKPTPPVVV-----VKK-----PTPTP 172
 DB 297 NGVFGRCPEAGKRAFDRSHPRAKTQHPFLGEBNTGTDVAVTASVAKKNEAQAASFTQG 356
 QY 173 PVVQOPAPVA-PVYTEAPFATG-----SGVNOFERYV-GATNPVRRFGTATVAGST 223
 DB 357 KLITETKAPAKESSEVAIRSDGNEDELPKSTIGIKYMPVPGA--VVAAYG-ANVDGN- 411
 QY 224 VTSQGMFSGRDGLINASNAGTVIOADHNM--GASIVIOHTNGFVSYIHIDKAOYKT 281
 DB 412 -RNDGINISVDEGPIITKAENGVIYSGSSIKELGNVLVRHDDGTVTYGNAALVKYQR 470
 QY 282 GDTVRTQRIASMKQPSGAAL---FEFRISRNGVYVDP 322
 DB 471 GQKVQROQTLAS--SGMTGRATRPQVHFVVRKDATPVNPATYLE 512

RESULT 24
 0926D2 PRELIMINARY; PRT; 512 AA.
 ID 0926D2
 AC 0926D2;
 DT 01-DEC-2001 (Tremblrel. 19, Created)
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
 DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
 DE Lipoprotein precursor.
 GN LPPB OR R01536 OR SMC02060.
 OS Rhizobium meliloti (Sinorhizobium meliloti).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Sinorhizobium.
 OC NCBI_TaxID=382;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1021;
 RX MEDLINE=21396507; PubMed=11481430;
 RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Bactut J.,
 RA Boistard P., Becker A., Boutry M., Cadieu B., Dreano S., Gloux S.,
 RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaire V., Maury D.,
 RA Pohl T., Portetelie D., Puenhler A., Purnelle B., Rampegger U.,
 RA Renard C., Thebault P., Vandenbol M., Weidner S., Galibert F.,
 RT "Analysis of the chromosome sequence of the legume symbiont
 RT Sinorhizobium meliloti strain 1021.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
 DR EMBL; AL591787; CAC46115.1; -.
 DR InterPro; IPR002482; LysM.
 DR InterPro; IPR002886; Peptidase_M37.
 DR Pfam; PF01476; LysM; 2.
 DR Pfam; PF01551; Peptidase_M37; 1.
 KW Complete proteome.
 SQ SEQUENCE 512 AA; 52003 MW; 6BBADD3290125D0 CRC64;

Query Match 10.4%; Score 174; DB 16; Length 512;
 Best Local Similarity 24.0%; Pred. No. 0.00012;
 Matches 79; Conservative 60; Mismatches 122; Indels 68; Gaps 12;
 QY 37 PTYNSTSGSHRTSGS-----GGLAIGSQVITDSQ----- 67

Db 209 PTEGTASAKAAAGSADVEGKKRPSPLPTDREVALIPGOSQREKSESRDVAAG 268
Qy 68 -----GYPNRQVQGPVSKIAQRYGLNMEIGHINNINSYITTYGOWLTMSGD 119
Db 269 KLASAGGGGNGATVTPGDLNRIKANGVSAVALQANGL-STEARIRIQKXINTI9A- 326
Qy 120 LKAKRENSISGCVTAHTPPSPVAVOSSRPVQOHPAVOKPTPPVVVKKPTPTPVQOPA 179
Db 327 -----SAKTPADVAVTASVAKKNEAQAASTEQKLT-ITAPAKESVS 371
Qy 180 PVAPVTEAPATGSSGVMOFRPVV-GATNPVRRFGTAVAGSTVTSNGKMFSGRDDL 238
Db 372 EVAIRSDGNEDLPKSTGIGKRWVRGA--VVAAYG-ANVDGN--RNDGINISVPRGTP 425
Qy 239 INASNACTVIOADHNM--GASIVIOHTNGFVSSYIHIKAQVKTGDTVRTGORIASMKN 296
Db 426 IKAENGVVIYSGSLKELGNAVLVRHDDGTIVYGNAAELIKVORGKVCRGQTLAS--S 483
Qy 297 OPSGAL--FEFRISRGVVDPLTVLK 322
Db 484 GMTGRATRPQVHFVRKNAATPVNPATYLE 512

RESULT 25

Q8RNB5 PRELIMINARY; PRT; 392 AA.
AC O8RNB5.
DT 01-JUN-2002 (TREMREL. 21. Created)
DT 01-JUN-2002 (TREMREL. 21. Last sequence update)
DT 01-JUN-2002 (TREMREL. 21. Last annotation update)
DE NLPD.
GN NLPD.
OS Bartonella henselae (Rochalimaea henselae).
OC Bacteria; Proteobacteria; alpha subdivision. Rhizobiaceae group;
OC Bartonellaceae; Bartonella.
OX NCBI_TaxID=38323;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HOUSTON-1;
RA Fiskus W., Padmalayam I., Baumstark B.;
RT "Molecular cloning, sequencing, and characterization of the NLPD/lppB
homolog in Bartonella henselae."
RL Submitted (FEB-2002) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF484425; AAL89761.1; -
SQ SEQUENCE 392 AA; 42635 MW; CF71E6B91FDA206C CRC64;

Query Match 10.3%; Score 172.5; DB 2; Length 392;
Best Local Similarity 24.0%; Pred. No. 0.00011;
Matches 79; Conservative 49; Mismatches 130; Indels 71; Gaps 12;

Qy 35 SKPTVNSTSGSGSHRTSGGGLAIG---SQVITDSQGVNPNRYOVKODPTVSKINORVGL 90
Db 93 NSPQDGTSSPNS-RIMGTTPRNIGTLSRSQWRNDPLFRQNSYIVQTGDTLISARQGV 151
Qy 91 WMRREIGHINNINSSTIYTGOWLTMSGD-----KXER 125
Db 152 SVEALIKVNGIRSN-SITIGVLMIPSGRTAETSVNRNDKASQSLSSQSSASISHK 210
Qy 126 SISGCVTAHTPPSPVAVOSSRPVQOHPAVO-----KPTPPVVVKKPTPTPVQOP 178
Db 211 KYSSSTEKAPITPKPFA-QINNSNGQSSAQMSLNTERGVLDITVMKNDKONGVTPQ----- 264
Qy 179 APVAPVTEAPATGSSGVMOFRPVVGAITNPVRRFGTAVAGSTVTSNGKMFSGRDDL 238
Db 265 -----ATGISKRWVR--RGRLLSFGQKR--GTTMRGIDIAVPESS 304
Qy 239 INASNACTVIOADHNM--GASIVIOHTNGFVSSYIHIKAQVKTGDTVRTGORIASMKN 296
Db 305 VKAENGVVIYASDGLKEIGNVMMIRHENNITTYIGNSKLVVTVGORIRGDEIA--KS 362
Qy 297 OPSG--AALFEFRISRGVVDPLTVLK 322
Db 363 GVSQGVKTPRVYFEVRENSLPPVDPIKYLE 391

RESULT 26

Q9X7M8 PRELIMINARY; PRT; 312 AA.
AC Q9X7M8.
DT 01-NOV-1999 (TREMREL. 12. Created)
DT 01-NOV-1999 (TREMREL. 12. Last sequence update)
DT 01-JUN-2002 (TREMREL. 21. Last annotation update)
DE Putative secreted peptidase.
GN SC06773 OR SC685.22.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Oliver K., Harris D.;
RL Submitted (MAR-1999) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Bentley S.D.; Parkhill J.; Barrell B.G.; Rajandream M.A.;
RL Submitted (MAR-1999) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kleser H.M., Denapate D., Richner A., Cullum J.,
RA Kinasli H., Hopwood D.A.;
RT "A set of ordered cosmid and a detailed genetic and physical map for
the 8 Mb Streptomyces coelicolor A3(2) chromosome."
RL M.J. Microbiol. 21:77-96(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RA Bentley S.D., Chater K.F., Cerdano-Farraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleser H.,
RA Harper D., Bateman A., Brown S., Chandra J., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kleser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Ruter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
coelicolor A3(2)."
RL Nature 417:141-147(2002).
DR EMBL; AL049485; CAB39706.1; -
DR InterPro; IPR002482; LysM.
DR InterPro; IPR002886; Peptidase_M37.
DR Pfam; PF01476; LysM; 1.
DR Pfam; PF01551; Peptidase_M37; 1.
DR PRINTS; PR01574; TUBBYPROTEIN.
DR SMART; SMO0257; LysM; 1.
SQ SEQUENCE 312 AA; 31633 MW; 2D489B9B7C448206 CRC64;

Query Match 10.3%; Score 171.5; DB 16; Length 312;
Best Local Similarity 23.3%; Pred. No. 9.5e-05;
Matches 79; Conservative 43; Mismatches 118; Indels 99; Gaps 15;

Qy 46 GSHR-----TSGSGGLAIGSQVITDS-----QGP----- 70
Db 5 GKRRRTAMRLRTITAVAGTGALALPLMGAAANAAPASVSEQAVQSPAAKAKAA 64
Qy 71 -----NKYOVQGDVSKIAQRYGLNMEIGHINNINSYITTYGOWLTMSGDLYRE 124
Db 65 EKNSDSRTVYKSSDYSLADEQDV-----GKKKUYADN-----R 102
Qy 125 RSISGCVTAHTPPSPVAV--QSSRPVQOHPAVOKPTPPVVVKKPTPTPVQOPAPVA 182
Db 103 EAVSDPSLHPGKLSIDGAAKPSAPSSAQSKFA-----OKSAEKPAEKTAAPFA 155


```

QY 183 PPVTAPATGSSGWMQFRYPVGTATNPVRRFGATATVAGSTV-----TSNGMFSG----- 233
D 156 QKSTDAKASSSDTQSG---STGTTSGY-----TSPVAGGTGVPYHQSGSMSSGVTHTG 207
QY 234 -----RUGDLINASNAGTIVQAD-HNMDGASIVQHTNGFVSSTYHKKDAQVKTGDTVRT 287
D 208 TDFVVPPTGTSIKAAAGATVVSAGGAYGNQVTVQLADGHYAQYAHLSLSVSAGOSVTA 267
QY 288 GORI--ASMKNQPSGAL-PEFRISRN-GVYVDELTVLK 322
D 268 GQVGLSGATGATGVTGPHLHFIRITTPYGSIDIDLAFUR 306

RESULT 27
Q8R6U3 PRELIMINARY; PRT; 449 AA.
AC Q8R6U3;
DT 01-JUN-2002 (TEMBLrel. 21, Created)
DT 01-JUN-2002 (TEMBLrel. 21, Last sequence update)
DE Membrane proteins related to metalloendopeptidases.
GN NPDP OR TTE2691.
OC Thermoanaerobacter tengcongensis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Thermoanaerobacteriales; Thermoanaerobacteriaceae; Thermoanaerobacter.
OX NCBI_TaxId=119072;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MB4T / JCM1007;
RX MEDLINE=21992816; Pubmed=11997336;
RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA Tan H., Chen R., Wang J., Yu J., Yang H.;
RT "A complete sequence of T. tengcongensis genome.";
RU Genome Res. 12:689-700(2002).
DR EMBL; AE013208; AAM25810.1; -.
SQ COMPLETE PROTEOME.
KW SEQUENCE 449 AA; 50731 MW; EA7A06CB6CD151F0 CRC64;

Query Match 10.2%; Score 171; DB 16; Length 449;
Best Local Similarity 27.6%; Pred. No. 0.00017;
Matches 74; Conservative 42; Mismatches 120; Indels 32; Gaps 12;

QY 73 YOYKQGDYTSKIAQRYGLANREIGHN-----NLNSYTYITGQWLTLMGSDLLVRRRSI 127
D 193 YEKENDSLMSIAREHMYIQDILKLPGLTENLKPGQIITLSKEVPLV---VTEKEY 249
QY 128 S---SGVNTAHT-PSPVAVQSSRPVQOHPAVOKPTPPVNVVKKPTPTPPVQOPAPVA 182
D 250 TYKBEIPEFTKFTKDDKLFVNQSKVLVEGKGL-KEIKAVIVISHNGVEVVRDIEKERVLK 308
QY 183 PPVTAPATGSSGWM-----QFRYPVGTATNPVRRFGATATVAGSTVTSNGMFSGDCD 237
D 309 EPISKI-VAAGSRVSVYATGYSPARGT--ITSRGPWGGFHT---GVDAIARYGS 361
QY 238 LINAASAGTIVQADHNMD-GASIVQHTNGFVSSTYHKKDAQVKTGDTVRTGRIASMKN 296
D 362 PTAASDGGIVIFKMGESGYLVAKIDHANGVTVYIGHASKLVAKVGDKVEKGCKIALV-- 419
QY 297 QPSGAL--FEFRISRNGVYVDELTVL 321
D 420 GATGHATGPHVHFVRKNGVPIIDMRFL 447

RESULT 28
Q8X3L6 PRELIMINARY; PRT; 223 AA.
AC Q8X3L6;
DT 01-MAR-2002 (TEMBLrel. 20, Created)
DT 01-MAR-2002 (TEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
DE Putative lipoprotein.

```

```

GN 24203.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxId=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; Pubmed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lam A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RU Nature 409:529-533(2001).
DR EMBL; AE005516; AAG57994.1; -.
DR InterPro; IPR002482; LysM.
DR InterPro; IPR002886; Peptidase_M37.
DR Pfam; PF01476; LysM; 1.
DR Pfam; PF01551; Peptidase_M37; 1.
DR SMART; SM00257; LysM; 1.
KW Lipoprotein; Complete proteome.
SQ SEQUENCE 223 AA; 23472 MW; 774D23D63C9C81C3 CRC64;

Query Match 10.1%; Score 169; DB 16; Length 223;
Best Local Similarity 24.0%; Pred. No. 9.3e-05;
Matches 63; Conservative 35; Mismatches 79; Indels 86; Gaps 8;

QY 29 ILAAGSKPTYNSTSGSGSHRTSGSGGAIAGSVITDSQGVPNRYQYKQDYSKIAQRY 88
D 9 ILAAGC-----SGSKSSDITGYSG-----SVYTVKRGDTLYRISRTT 44
QY 89 GLANREIGHNINLNSYTYITGQWLTLMGSDLK---VREKSISSGVNTAH-TSPVAVQ 143
D 45 GTSVKELARLNGISPPYITVEGQKTLK--CGAKSSSTRKSTAKSTYKTASVTPSSAVPK 102
QY 144 SSRPVGQGHAVQKPPPPVNVVKKPTPTPPVQOPAPVAPVTEAPATSSGWMQFRYP 203
D 103 SSMPVGQRCWLMPTTGKVTM-----PSTADGG----- 131
QY 204 VGATNPVRRFGATATVAGSTVTSNGMFSGRDGLINASNAGTIVQADHNMDGAS--IVI 261
D 132 -----NKGIDISAPRGTPPIYAAGAGKVVYVGNQLRGIGNIMI 169
QY 262 QHTNGFVSSTYHKKDAQVKTGDT 284
D 170 KHSEDIYTAVAHNDTMLVNNNGS 192

RESULT 29
ID 083190 PRELIMINARY; PRT; 371 AA.
AC 083190;
DT 01-NOV-1998 (TEMBLrel. 08, Created)
DT 01-NOV-1998 (TEMBLrel. 08, Last sequence update)
DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
DE Hypothetical protein TP0155.
GN TP0155.
OS Treponema pallidum.
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxId=160;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NICHOLS;
RX MEDLINE=98332770; Pubmed=9665876;
RA Fraser C.M., Norris S.J., Weinstein G.M., White O., Sutton G.G.,
RA Dodson R., Gwin M., Hickey E.K., Clayton R., Ketchum K.A.,
RA Sodergren E., Hardham J.M., McLeod M.P., Salberg S., Peterson J.,
RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utecherback T.,
RA McDonald L., Attiach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,

```

RA "Venter J.C.;
 RT "Complete genome sequence of *Treponema pallidum*, the syphilis
 RL *spirochete*,"
 DR EMBL: AF001200; AAC65145.1; -
 DR EMBL: AF01551; AAC65145.1; -
 DR EMBL: AF01551; AAC65145.1; -
 DR InterPro: IPR002482; LysM
 DR InterPro: IPR002486; Peptidase_M37.
 DR Pfam: PF01476; LysM; 2
 DR Pfam: PF01551; Peptidase_M37; 1.
 DR SMART: SM00257; LysM; 2.
 DR Hypothetical protein; Complete proteome.
 KW SEQUENCE 371 AA; 40664 MW; 8880055E3A82A8 CRC64;

Query Match 10.0%; Score 167; DB 16; Length 371;
 Best Local Similarity 24.9%; Pred. No. 0.00025;
 Matches 72; Conservative 47; Mismatches 122; Indels 48; Gaps 11;

QY 73 YGVKQDTVSKIAQRYGLNMRREIGHINNLSSTIYTGOMLTWSGD---LKVRESISS 129
 DB 87 YEVREGDVGSRIQRYDISDPAISLNKLRSTRALQVQGLKIPSDVIGILTVKNGDPTFS 146
 QY 130 GVNTAHTPSPVAVQSSRPVQOHFAVOKFTPPVV-VVKKP-----TPTP----- 172
 DB 147 SIAMAH-----QISLERLVLTNPSSSKESPVRITIVSPFYNASARESCEVFPFSSAKQ 201
 QY 173 -----PVVQ--QPAPVAPVTEAPFATGSSGVQ-----PPYPGATNPVVRFPQ-- 215
 DB 202 WRENTSPDAVQPLQPA-----RVLPFLPGHLSARALQELINGDLFAPLRSRYVSRYGWR 257
 QY 216 TATVAGSVTVSNGWMPFGSDGDLINASAGVIGADNH-MGASIVIOHTNGFVSSYTHI 274
 DB 258 SDPTGASFSNGLDMSRRGTTPVYSAIGLVIRVGVGAIVGNLIVGHAGVQTLVYHL 317
 QY 275 KDAQVKTGDIYRTGQRIASM-KNOPSICALPEFRISNGVYVDPITVIX 322
 DB 318 QTVLVSAGTRVTSARKIGLTKRGTSGPHLFTIKNGSAINPTSLIK 366

RESULT 30
 Q9KJW8 PRELIMINARY; PRT; 401 AA.

AC Q9KJW8; 01-OCT-2000 (TREMblrel. 15; Created)
 DT 01-OCT-2000 (TREMblrel. 15; Last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19; Last annotation update)
 DE 43 kDa antigen.
 OS *Bartonella bacilliformis*.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Bartonellaceae; Bartonella.
 CX NCBI_TaxID=774;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=KCS84;
 RX MEDLINE=20407302; PubMed=10948113;
 RA Padmalayam I, Kelly T, Baumstark B, Massung R;
 RT "Molecular cloning, sequencing, expression, and characterization of an
 RT immunogenic 43-kDa outer lipoprotein of *Bartonella bacilliformis* that
 RT has homology to NlpD/LppD.";
 RL Infect. Immun. 68:4972-4979(2000).
 DR EMBL: AF157831; AAF60360.1; -
 DR InterPro: IPR002482; LysM.
 DR InterPro: IPR002886; Peptidase_M37.
 DR Pfam: PF01476; LysM; 1.
 DR Pfam: PF01551; Peptidase_M37; 1.
 DR SMART: SM00257; LysM; 1.
 SQ SEQUENCE 401 AA; 43418 MW; C5ABF0E8575FBE CRC64;

Query Match 9.9%; Score 165.5; DB 2; Length 401;
 Best Local Similarity 25.0%; Pred. No. 0.00036;
 Matches 83; Conservative 51; Mismatches 141; Indels 57; Gaps 14;
 QY 31 AGCASKPTYN-STSGSGS-----HRTSGSG-----GLAIGSVITDSOGVNNRYOV 75

DB 86 SCOHNASVNFPSHOEAGSSFRRENKVPDGRIMGSPONIGTLRSNVNPNPISIFOKGSYIV 145
 QY 76 KQGDVTSKIAQRYGLNMRREIGHINNLSSTIYTGOMLTWSGDLKVRERSISSGVNTAH 135
 DB 146 QSGDTLFSIAQKGISVESIKLVANGMDN-AIYIGKLVIPDKQVTVTSNAONNAVGSSST 204
 QY 136 TSPPVAVQSS--SRPPVQOHFAVOKP--FPVAVVKKFTPTPPVQOPAPVAPVTEAPF 190
 DB 205 TTS--STQSKGMSFVKKEBISASRSVNTAPAIIN---TPDVKN---VSKOMTSSKS 255
 QY 191 ATGSS-----GVMQFPYVATNPVVRFGTATVAGSVTVSNGWMPFGSD 235
 DB 256 ASNSKTVKDSNITVPAQFVSKMRPVRG-----RLSHVQOKGKITNRGIDINVE 310
 QY 236 GDLINMSAGTVIGADNHND--GASIVIOHTNGFVSSYTHIKQAVKGTGDIYRTGRIAS 293
 DB 311 GSSVKAENGVVIVYASDGLKELGVNVMIRHEDNIIITVGHNSKLTYSRGQVXRGDEIA 369
 QY 294 MNQPSGAL--FEFRISNGVYVDPITVIX 322
 DB 370 -KSGVSGDATTPRVYFEIRONSVPVNPAYOLE 400

RESULT 31
 Q8Y4E2 PRELIMINARY; PRT; 436 AA.

AC Q8Y4E2; 01-MAR-2002 (TREMblrel. 20; Created)
 DT 01-MAR-2002 (TREMblrel. 20; Last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21; Last annotation update)
 DE Hypothetical protein lmo2504.
 GN LMO2504.
 OS *Listeria monocytogenes*.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
 OC Listeriaceae; Listeria.
 CX NCBI_TaxID=1639;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=EGD / SERVAR 1/2A;
 RX MEDLINE=21537279; PubMed=11679669;
 RA Glaser P, Frangeul L, Buchrieser C, Ruenick C, Amend A,
 RA Baquero F, Berche P, Bloeker H, Brand P, Chakraborty T,
 RA Charbit A, Chetouani F, Couve E, de Daruvar A, Denoux P,
 RA Domian E, Dominguez-Bernal G, Duchaud E, Duran L, Duseigneur O,
 RA Entian K-D, Feihl H, Garcia-del Portillo F, Garrido P,
 RA Gautier L, Goebel W, Gomez-Lopez N, Hahn T, Haut U, Jackson D,
 RA Jones L-M, Kaerst U, Kreft J, Kuhn M, Kunst F, Kutapkat G,
 RA Madueno E, Maitournam A, Mata Vicente J, Ng E, Nedjari H,
 RA Nordstedt G, Novella S, de Pablos B, Perez-Diaz J-C, Purcell R,
 RA Remmel B, Rose M, Schluter T, Simoes N, Tierrez A,
 RA Ramirez-Boland J-A, Voss H, Weiland J, Cossart P;
 RT "Comparative genomics of *Listeria species*,"
 RT Science 294:849-852(2001).
 DR EMBL: AL591983; CAD00582.1; -
 DR ListerList; LMO02504;
 DR InterPro: IPR002886; Peptidase_M37.
 DR InterPro: IPR002017; Spectrin.
 DR Pfam: PF01551; Peptidase_M37; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 436 AA; 47049 MW; 2D36887C1098E167 CRC64;

Query Match 9.5%; Score 158.5; DB 16; Length 436;
 Best Local Similarity 24.8%; Pred. No. 0.0013;
 Matches 66; Conservative 33; Mismatches 100; Indels 67; Gaps 11;

QY 76 KQGDVTSKIAQRYGLNMRREIGHINNLSSTIYTGOMLTWSGDLKVRERSISSGVNTAH 135
 DB 210 EKNDLVMAANKKQIT-----KSEQTLAS-----QALITDEKILASINAGK 254
 QY 136 TSPPVAVQSSRPVQOHFAVOKFTPPVAVVKKFTPTPPVQOPAPVAPVTEAPATGSS 195
 DB 255 AKQEAATKAEKEMQEAALAAASAKSAVVK-----QPSSSNEATE-----TVSS 300

QY 196 GVMQFRYPVGA-----TNPVRRF-----GTATVAGSTVTSNGMFGSGDGLINAS 242
 DB 301 GGGGFIKIPAGGILITSGSERNTPTGKYESHKGGDIDAGGIVT-----VSAA 347
 QY 243 NAGTVIOADNMDCAS-----IVIOHTNGFVSSYIHKDAQV--TGDTRTGRIAS 293
 DB 348 ASGTIVPFGAGSGSGFGGYGVKIDHNGFQTLGYHMRAGSLKVTVGQGVSGGPIGI 407
 QY 294 M--KNQSSGAALFFFRISRNGVYDP 317
 DB 408 MGSTGSGTGOHL-HFEIRHNGVPPNDP 432

RESULT 32

QY 08RG41 PRELIMINARY; PRT; 321 AA.
 AC 08RG41;
 DT 01-JUN-2002 (Tremblrel, 21, Created)
 DT 01-JUN-2002 (Tremblrel, 21, Last sequence update)
 DT 01-JUN-2002 (Tremblrel, 21, Last annotation update)
 DE Cell wall endopeptidase family M23/M37.
 GN PR0477.
 OS Fusobacterium nucleatum (subsp. nucleatum).
 OC Bacteria; Fusobacteria; Fusobacterium.
 OX NCBI_Taxid=76856;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 25586;
 RX MEDLINE=21886394; Pubmed=11889109;
 RA Kapatcharya A., Barton A., Gardner W., Grechkin G., Zhu L.,
 RA Vasiela O., Chu L., Kogan Y., Chaga O., Goldsman B., Bernal A.,
 RA Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R.,
 RA Fomstein M., Kyrides N., Overbeck R.,
 RA "Genome sequence and analysis of the oral bacterium Fusobacterium
 RT nucleatum strain ATCC 25586."
 RL J. Bacteriol. 184:2005-2018(2002).
 DR EMBL: AE010559; AAL94673.1; -
 KW Complete proteome.
 SQ SEQUENCE 321 AA; 35455 MW; 0E4158B02AA52DD9 CRC64;

Query Match 9.5%; Score 158; DB 16; Length 321;

Best Local Similarity 24.9%; Pred. No. 0.00093;
 Matches 85; Conservative 49; Mismatches 125; Indels 82; Gaps 18;

QY 5 IAINSONQKPKIKLIGVITTCILAGCASKPTTNTSGSGSHRTSGSGGLAIGSVYT 64
 DB 40 IITYVQCKDVTVOVAKKFGVPEITMI-----NNQFAMDNKLKVG-EVLT 83
 QY 65 DSGQVPRRYGVKODYTSKIAQRYGLMWRREIGHINNINSSYITYGQWLTMSGDLK- 121
 DB 84 PPSIDGLYTLQKEMAKVAKKYGAVVDIGDNNINPK-KLAGTTLFLKGVTLKKYK 142
 QY 122 -VERSISSGVNTHTSPVAVQSSRPVQOHPAVOKPTPPVAVVKKPTPPVQOPAP 180
 DB 143 EVEGRRL-----AAQAKKE-QKKEAQQKGG-----KKGSGAP----- 176
 QY 181 VAPPVT-----EAPFATGSGVMQFRYPV--GATNPVRRFETATVAGSTVTSNGMP 231
 DB 177 --PPDTGGDGGAGPASYSGE--FAFPYVAGITSPFGNRY--HPLAKYIHLHTV-- 227
 QY 232 SGRGDLI-----NASNAGTVIOADNMDCAS--IYIOHTNGFVSSYIHKDAQVETGD 283
 DB 228 ----DLVAKYVPLRASKAGVVTAG--NMSGYKIIIIKHNGYETRAHLSTVSTNVE 281
 QY 284 TVRTGRIASMKN--QPSGAALFFFRISRNGVYDPPLTVLK 322
 DB 282 HVNKGDLIGTKNGSGRTTGAIL-HFEIRHNGVPPNDPKYLO 321

RESULT 33

Q9RMW21

ID Q9RMW21 PRELIMINARY; PRT; 375 AA.
 AC Q9RMW21;
 DT 01-MAY-2000 (Tremblrel, 13, Created)
 DT 01-MAY-2000 (Tremblrel, 13, Last sequence update)
 DT 01-JUN-2002 (Tremblrel, 21, Last annotation update)
 DE Lipoprotein, putative.
 GN DR0848.
 OS Deinococcus radiodurans.
 OC Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;
 OC Deinococcaceae; Deinococcus.
 OX NCBI_Taxid=1299;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=R1;
 RX MEDLINE=20036896; Pubmed=10567266;
 RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
 RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
 RA Motil K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
 RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zaleski C.,
 RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
 RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
 RA Fraser C.M.;
 RT "Genome sequence of the radioresistant bacterium Deinococcus
 RT radiodurans R1."
 RL Science 286:1571-1577(1999).
 DR EMBL: AE001939; AAF10427.1; -
 DR TIGR: DR0848;
 DR InterPro: IPR002482; LysM.
 DR InterPro: IPR002886; Peptidase_M37.
 DR Pfam: PF01476; LysM; 1.
 DR Pfam: PF01551; Peptidase_M37; 1.
 DR SMART: SM00257; LysM; 1.
 KW Lipoprotein; Complete proteome.
 SQ SEQUENCE 375 AA; 39107 MW; 3FF9A242F97C00AB CRC64;

Query Match 9.5%; Score 158; DB 16; Length 375;

Best Local Similarity 21.7%; Pred. No. 0.0011;
 Matches 74; Conservative 36; Mismatches 105; Indels 126; Gaps 12;

QY 68 GVPNRVYVQKODYTSKIAQRYGLMWRREIGHINNINSSYITYGQWLTMSGDLKVRERSI 127
 DB 14 GVAGAVTVKKKGLDLYSLARGSGLTVDALMRLNGL-STPELRVGGVITL-----P 61
 QY 128 SSGVNTHTSPVAVQSSRP-PVQOHPAVOKPT--PPV-----VKKP--TPRP----- 173
 DB 62 GEGATPATPATPAAPPTATPAAPAPAPAVPTGLPVAANAVLKIPTITTPKALM 121
 QY 174 -----VQOPAPVAPPVTE-----APFATGSGVMQFRYPVAGATNPV- 211
 DB 122 GDGFALRLSGEQANVTVRRPSELGEDVROPNELRLMSGQYIVPGRVVLGKTTPIY 181
 QY 212 -----RR----- 213
 DB 182 EYRLGSDVVRGQIPVGDLGQRVQHLNLPISIKYLOPAREABDAVEAOAYORFTPOOMQ 241
 QY 214 --FGTATVAGSTVTSN-----GMWFSGRDGLINASNAGTVIOAD-HN 253
 DB 242 QPFSALASGKATSSSGQRTYVAGAKVAYHGTIDYPAASGTAVLAINDSIVIAGRYP 301
 QY 254 MDGASIVIOHTNGFVSSYIHKDAQVETGDTVRTGRIASM 294
 DB 302 VRGGLVVIDHGAQVLSLYFHQSKVTAKVGGQIKRQGVGEV 342

RESULT 34

Q927Y9

ID Q927Y9 PRELIMINARY; PRT; 437 AA.
 AC Q927Y9;
 DT 01-DEC-2001 (Tremblrel, 19, Created)
 DT 01-DEC-2001 (Tremblrel, 19, Last sequence update)
 DT 01-MAR-2002 (Tremblrel, 20, Last annotation update)
 DE Hypothetical protein lin2647.
 GN LIN2647.

```

OS Listeria innocua.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Listeriaceae; Listeria.
NCBI_TaxID=1642;
RN
RP SEQUENCE FROM N.A.
RC STRAIN-CLIP_11252 / SEROVAR 6A;
RX PubMed-11679669;
RA Glaeser F., Fraenkel L., Buchrieser C., Rusniok C., Amend A.,
RA Baquero F., Berche P., Blocker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetoui F., Cove E., de Daruvar A., Denoux P.,
RA Domian E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entian K.-D., Fath H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kreft J., Kunz M., Kuntz F., Kurapkat G.,
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nardone G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schluter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RA "Comparative genomics of Listeria species";
RL Science 294:849-852(2001).
DR EMBL; AL596173; CAC97874.1; -.
DR ListList; L1N02647; -.
DR InterPro; IPR002886; Peptidase_M37.
DR Pfam; PF01551; Peptidase_M37; 1.
DR Hypothetical protein; Complete proteome.
SQ SEQUENCE 437 AA; 47301 MW; F3PD028682428C57 CRC64;

Query March 9.3%; Score 155; DB 16; Length 437;
Best Local Similarity 23.3%; Pred. No. 0.0023;
Matches 62; Conservative 35; Mismatches 10; Indels 66; Gaps 10;

QY 76 KQGTVAKIAIQRVGLNREIGHNNNSYITTCOMITLMSGDLKVRERISGVNNAH 135
DB 210 EKNLIVMALANKKDLT-----KSEQTLITNE-----QGLSDQCKKLAIAGEK 254
QY 136 TPSPVAVOVSRRPVQGHFAVQKPTPPVVVKKPTPPVVOQPAFVAPVTEAPATGSS 195
DB 255 AKQENAIKAAEKRRQENAAAKNAT-----VAAQPPSSVTSSAGTATDTVSS 301
QY 196 GVMQFRYPVGA-----TNPVVRP-----GTATVAGSTVTSNGMWSGRDGLINAS 242
DB 302 GCGGFIKPAAGMLTSGFSERTNTPVKGKESHKQDIAAGGIT-----VSAA 348
QY 243 NAGTVIQADHNMMDGAS-----IVIOHTNGFVSSYIHKDAQVK--TQDTVRTGORIAS 293
DB 349 ASGRVVFSGFATGSGFGGYYVVKIDHNGFQTLYAHMRAGSLKVTGQGVSGQPIGI 408
QY 294 M--KNQPSGALFEPRISRNGVYVDP 317
DB 409 MGTGQSTGQHL-HREIHKNGIPVDP 433

RESULT 35
OS2606 PRELIMINARY; PRT; 176 AA.
ID OS2606;
AC 01-JUN-1998 (TREMBlrel. 06, Created)
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Novel 11poprotein N1PD (Fragment).
GN N1PD.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC Ralstonia.
NCBI_TaxID=305;
RN
RP SEQUENCE FROM N.A.
RC STRAIN=AW1;
RC MEDLINE=98294041; PubMed=9632252;
RA Flavier A.B., Schell M.A., Denny T.P.;
RA "An Rpos (sigmas) homologue regulates acylhomoserine lactone-dependent

```

```

RT autoinduction in Ralstonia solanacearum.";
RL Mol. Microbiol. 28:475-486(1998).
DR EMBL; AF042351; AAC38576.1; -.
DR InterPro; IPR002886; Peptidase_M37.
DR Pfam; PF01551; Peptidase_M37; 1.
DR Lipoprotein.
FT NON TER.
SQ SEQUENCE 176 AA; 18450 MW; EBB5214BAEA6300 CRC64;

Query March 9.2%; Score 153.5; DB 2; Length 176;
Best Local Similarity 28.8%; Pred. No. 0.0009;
Matches 51; Conservative 26; Mismatches 79; Indels 21; Gaps 6;

QY 153 PAVQKPTPPVYVKKPTPTP--VVOQPAFVAPVTEAPATGSS-----GVMQFRYPV 205
DB 10 FNAADANTNGVVV-----TPVRPGNTTSDPITRAVTSPPVVASGASGATGATLAMP-- 63
QY 206 ATNPPVRPPTATVAGSTVTSNGMWSGRDGLINASNACTVIOADHMD--GASIVYQHT 264
DB 64 AHQGVIGRF-----DDKANKGIDIGKKRQDPVSAADSKVHVGLRGVGNLVITKHN 116
QY 265 NGFVSSYIHKDAQVKTDGTVRTGQRIASMKNOPSGAALFEPRISRNGVYVDP 321
DB 117 DTELTAYGNNDKVLVTESQTVKKQKLAEMGSTDADRVKLFVRRNGKVPDWRFL 173

RESULT 36
OS2606 PRELIMINARY; PRT; 228 AA.
ID OS2606;
AC 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Putative peptidase.
GN SC03168 OR SCE94.19C.
OS Streptomyces coelicolor.
OC Actinobacteria; Actinobacteriales; Streptomyces.
OC Actinomycetales; Streptomyces; Streptomyces.
NCBI_TaxID=1902;
RN
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RC Oliver K., Harris D.;
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RC Bentley S.D.; Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
[3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RC Bentley S.D.; James K.D., Harris D.E., Quail M.A., Kleser H.,
RC Thomson N.R., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RC Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RC Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RC Huang C.-H., Kleser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RC Rabinowitsch E., Rajandream M.A., Rutherford K., Ruter S.,
RC Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RC Warren T., Wietzorek A., Woodward J., Barrell B.G., Parkhill J.,
RC Hopwood D.A.;
RC "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RT Nature 417:141-147(2002).

```

DR EMBL, AL049628, CAB40868.1; -
DR InterPro; IPR002886; Peptidase_M37.
DR Pfam; PF01551; Peptidase_M37; 1.
SQ SEQUENCE 228 AA; 22817 MW; 2263DF3167B0D279 CRC64;

Query Match 9.2%; Score 153; DB 16; Length 228;
Best Local Similarity 37.7%; Pred. No. 0.0014;
Matches 46; Conservative 16; Mismatches 42; Indels 18; Gaps 6;
QY 219 VAGSTVSNQMW---FSGRD-----GDILNANAGTYIADHN-----MGASIVIOHT 264
DB 101 LSASFQNGGMMAHSGQDPAVIGTIVVAHGGTIVKAGGNGAGDGPAYGMAIVIKHG 160
QY 265 NGFVSSYIHKDAQVKTGDTYRTGQRIASMKR--QPSGAL--FEFRISRN--GVYVDPPLTV 320
DB 161 NGTSSQYAHLSRIWVKIGQIVKTQSLAKSGNTGNSGPHLFEIRTPNPGNSAVDPVAF 220
QY 321 LK 322
DB 221 LR 222

RESULT 37
Q9X6S4 PRELIMINARY; PRT; 203 AA.
AC Q9X6S4;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Lipoprotein NlpD (Fragment).
GN NLPD.
OS Vibrio parahaemolyticus.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_Taxid=670;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BB22;
RA McCarter L.L.;
RT "RNA polymerase sigma factor Rpos of Vibrio parahaemolyticus";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF144608; AAD33932.1; -
DR InterPro; IPR002886; Peptidase_M37.
DR Pfam; PF01551; Peptidase_M37; 1.
FT NON TER 1
SQ SEQUENCE 203 AA; 21915 MW; 33EB174EADFA2360 CRC64;

Query Match 9.1%; Score 152.5; DB 2; Length 203;
Best Local Similarity 26.4%; Pred. No. 0.0013;
Matches 56; Conservative 34; Mismatches 75; Indels 47; Gaps 9;
QY 137 PSPVAVQSSRPVQGHVAVOKP---TPPVVVVK---KPPPTPVVQGPAP----- 180
DB 3 PVVVVVVTTTP---PPVAKFTTTTKPVVSSKSSQKPTTTPKVAQKEBPKEVEQTKA 57
QY 181 -----VAPVTEAFATGSSGVMQFYPVAGATPVVRRGATVAVASTVTSN 227
DB 58 KEYVSGKDNQHTKPKPTT---TVQNDKYSKMLMP---TKGRVAKNS---AEQ--GNK 106
QY 228 GMMFSGDGDILNANAGTYIADHNMDGS--IVIQHTNGFVSSYIHKDAQVKTGDTV 285
DB 107 GIDIAQGRGQPIVSTAAGTIVVYSGNALRGYNLIIYKHNIVYLSAVAHNDGLVTEGQSV 166
QY 286 RTGORIASMKRQPSGALFEFRISRNGVYVDP 317
DB 167 KSGQKLTMGSSGAKSVKLFHFRYQGSVNP 198

RESULT 38
Q8YRUO PRELIMINARY; PRT; 760 AA.
AC Q8YRUO;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)

DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Hypothetical protein Alr3353.
GN ALR3353.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_Taxid=1036590;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneke T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Res. 8:205-213(2001).
DR EMBL; AP003592; BAB75052.1; -
DR InterPro; IPR002482; LysM.
DR InterPro; IPR002886; Peptidase_M37.
DR Pfam; PF01476; LysM; 1.
DR Pfam; PF01551; Peptidase_M37; 1.
DR SMART; SM00257; LysM; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 760 AA; 80569 MW; FB62A0DDFCA573B CRC64;

Query Match 9.1%; Score 152; DB 16; Length 760;
Best Local Similarity 20.1%; Pred. No. 0.0079;
Matches 95; Conservative 41; Mismatches 112; Indels 224; Gaps 18;
QY 61 QVITD--SQGVNRYOVKQDQVSKIAQRYGIANREIHNINLNSYTYTGWTLWMSGD 119
DB 294 QPLTETSASNTAYEVKPGDTLAAISRVNTSVAEIVKVNLSNPDLKLSQQLIIS-- 351
QY 120 LKVRNSISGCVNTATPSS-----PVAQSSRP--PVQGHVAVOK-- 157
DB 352 -----AIDSSSTIAQPAIISNRVQPGTPVPVNPVDTARVNPFLPVAQPPSIANN 405
QY 158 -----PTP-----PV-----PVVQSSRP--PVQGHVAVOK-- 168
DB 406 SVPTVFTPGADSETPTDTIVPLESASAPAEIQGVGNVPIPAFLIQRPOQPKCAAR 465
QY 169 -----TPPPVVOQ----- 177
DB 466 AKGDBRLSLQAEIQRLQAKRYDQGTGTPPVVANQNNNTAIPVTSNNPNTATRISR 525
QY 178 ----PAVAPPTVAPATSSGVMQFR--YPVGA-----NPVRRFGTATV 219
DB 526 QOEIAVPIAVPTPLAPDSNHPVKPQGRATLPVNEALNPEFLPNNAPONSVRRVATPPV 585
QY 220 -----AGSTV-----TSNG 228
DB 586 RLNASESLGMRKGTVPSPKLPPLAANDQVLPQALIDESVPPSDSTTAFIWPANGVLITGY 645
QY 229 MWFSGR-----DGDILNANAGTYIADHNMDGS--IVIQHTNGFVSSYIHKD 276
DB 646 GMMFSGDGDILNANAGTYIADHNMDGS--IVIQHTNGFVSSYIHKD 285
QY 277 AQVKTGDTYRTGQRIASMKRQPSGALF-----EFRISRNGV--YVDPPLTVL 321
DB 706 ILVQAGQGVVQGOQIAM-----GSTGFTSGPTHFRIHTKGALNPIMLV 752

RESULT 39
Q9FTX9 PRELIMINARY; PRT; 176 AA.
AC Q9FTX9;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE Lipoprotein (Fragment).
GN NLPD.
OS Escherichia coli.

```
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OX Escherichia.
NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CECT073;
RX MEDLINE=20485587; PubMed=11029456;
RA Culham D.E., Wood J.M., Reference collection group B2- and uropathogen-
RT associated polymorphism in the rpoS-mutS region of the E. coli
RL chromosome.
RL J. Bacteriol. 182:6272-6276 (2000).
DR EMBL; AF270457; AAC27176.1.
DR InterPro; IPR002886; Peptidase_M37.
DR Pfam; PF01551; Peptidase_M37; 1.
FT NON_TER
SQ SEQUENCE 176 AA; 18847 MW; 3CBF32F8582D6F1A CRC64;

Query Match 8.8%; Score 147.5; DB 2; Length 176;
Best Local Similarity 27.3%; Pred. No. 0.0024;
Matches 54; Conservative 30; Mismatches 81; Indels 33; Gaps 8;

QY 126 SISSGVTAHTPPSPVAVQSSRPVQHPAVQKPTPPVVVKKPTPTPPVQOPAPVAPV 185
DB 7 SESSGEQANMLP-----NNKPTATVTPVTPVTPAST-TEDIVSTST----- 49
QY 186 TEAPFATGSSGVMPQRYPVGATNPVRRPGTATVAGSTVTSNGMFGSGDDGLINASVAG 245
DB 50 TSTPST-----WRWP--TEGKVIETFG-ASEG---NKGIDIASKGOAIIATADG 95
QY 246 TVIAGDHMDGAS--IVIOHNGFVSSYTHIKDAQVKTGDTYRTGORTASMKNOPSGAAL 303
DB 96 RVVYAGNLRGVGNLIIKHNDYISAYAHNDYMLVREQEVKAGOKIATMGSTGTSSTR 155
QY 304 PEFRIIRNGVYVDPITVLK 321
DB 156 LHFETRYKGSVNPLRYL 173

RESULT 40
Q9S4T2 PRELIMINARY; PRT; 247 AA.
AC Q9S4T2;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
DE Novel 11poprotein homolog.
GN NLPD.
OS Legionella pneumophila.
OC Bacteria; Proteobacteria; gamma subdivision; Legionellaceae group;
OC NCBI_TaxID=446;
CX [1]
RN SEQUENCE FROM N.A.
RC STRAIN=PHILADELPHIA-1;
RX MEDLINE=99369860; PubMed=10438758;
RA Hales T.M., Spuman H.A.;
RT "The Legionella pneumophila rpos gene is required for growth within
RL Acanthamoeba castellanii".
RL J. Bacteriol. 181:4879-4889 (1999).
DR EMBL; AF117715; AAD51395.1; -.
DR InterPro; IPR002482; LysM.
DR InterPro; IPR002886; Peptidase_M37.
DR Pfam; PF01476; LysM; 1.
DR Pfam; PF01551; Peptidase_M37; 1.
DR SMART; SM00257; LysM; 1.
KW Lipoprotein.
SQ SEQUENCE 247 AA; 27535 MW; 6C36631BD5BC9675 CRC64;

Query Match 8.7%; Score 145; DB 2; Length 247;
Best Local Similarity 24.3%; Pred. No. 0.0058;
Matches 78; Conservative 43; Mismatches 104; Indels 96; Gaps 15;
```

```
QY 16 KRIGLFGVITTCILAGCASKPTYNSTSGSHRTSGGGLAIGSVITDSQVNP-NRYQ 74
DB 7 KOLICIFFILSA---AGCGTR-----SDLA-----PVTELKMQPYSKYQ 42
QY 75 ---VKQDITVSKIAQRYGLNREIGHINNLSYTIYTGQWTLTWSGDKVRSISSG 130
DB 43 KPHTVRSGETLVAIAFYDIDYRRRLAINNHLRPPEYSLRVGVQVLSL-----KGIVPK 93
QY 131 VNTAHTPSPVAVQSSRPVQHPAVQKPTPPVVVKKPTPTPPVQOPAPVAPVTEAPF 190
DB 94 SNHFN-----SABRALH-----ITHKP-----KQKTIYSP---AYN 122
QY 191 ATGSSGVMPQRYPVGATNPVRRPGTATVAGSTVTSNGMFGSGDDGLINASVAGTIOA 250
DB 123 VRSSSGMLN-----PTGKAVTTFIPSQ--GKKGINIACNRGDKVRSASGVVAYA 171
QY 251 DHNMDGAS--IVIOHNGFVSSYTHIKDAQVKTGDTYRTGORTASMKNOPSGAALFE--- 305
DB 172 GSGLAGYGNLIIKHSNEYLTAGNNARNLVTBQRVNAGQVLAIE-----AGLIRSY 224
QY 306 ---FRISRNGVYVDPITVLK 322
DB 225 WGVHFEIRFRAQVPPVNPPLNYLQ 245
```

Search completed: July 8, 2003, 11:07:35
Job time : 85 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 8, 2003, 11:08:28 ; Search time 72 Seconds
(without alignments)
595.926 Million cell updates/sec

Title: US-10-018-706-2

Perfect score: 322
Sequence: 1 MTVTIAINSQNGKRIKRLGL.....LFEFRISRGVYVDPLTLVK 322

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 908470 seqs, 133250620 residues

Word size : 0

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database : A_Geneseq_101002.*

1: /SID82/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID82/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID82/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID82/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID82/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID82/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID82/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID82/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID82/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID82/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID82/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID82/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID82/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID82/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID82/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID82/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID82/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID82/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID82/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID82/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID82/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID82/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID82/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	322	100.0	322	22	AA820106	Moraxella catarrha
2	221	68.6	322	22	AA820107	Moraxella catarrha
3	9	2.8	968	22	AB861691	Drosophila melanog
4	8	2.5	223	22	AA890076	C glutamicum prote
5	8	2.5	557	20	AAW94997	YAK-1 related seri
6	7	2.2	13	18	AAW31599	Linker peptide use
7	7	2.2	13	23	AAU11448	Synthetic peptide
8	7	2.2	35	22	AA810936	Mouse mass protein
9	7	2.2	54	23	ABP04087	Mouse CSF protein
10	7	2.2	66	22	AAW83197	Human immune/haema

11	7	2.2	67	20	AAV41355	Human secreted pro
12	7	2.2	72	22	AAU61587	Proionbacterium
13	7	2.2	80	22	AAU50707	Proionbacterium
14	7	2.2	90	21	AAV96225	Arabidopsis ltrc1 c
15	7	2.2	101	23	AB855442	Lactococcus lactis
16	7	2.2	108	21	AA658104	Arabidopsis thalia
17	7	2.2	112	22	AB866134	Drosophila melanog
18	7	2.2	114	23	ABP00920	Human ORFX protein
19	7	2.2	126	22	ABG04226	Novel human diagno
20	7	2.2	136	22	AAO05078	Human polypeptide
21	7	2.2	138	21	AA658103	Arabidopsis thalia
22	7	2.2	147	22	AAO05875	Human polypeptide
23	7	2.2	157	22	AAE02227	Staphylococcus aur
24	7	2.2	159	22	AAU66705	Proionbacterium
25	7	2.2	159	22	AAU43362	Proionbacterium
26	7	2.2	179	23	ABP25478	Streptococcus poly
27	7	2.2	185	21	AAV75633	Neisseria gonorrhe
28	7	2.2	195	22	AB861524	Drosophila melanog
29	7	2.2	208	19	AAW71722	Arabidopsis leafy-
30	7	2.2	208	21	AAV54563	Amino acid sequenc
31	7	2.2	208	22	AAE09838	leafy cotyledon 1
32	7	2.2	213	22	AB859336	Drosophila melanog
33	7	2.2	251	22	AAU19650	Human novel extrac
34	7	2.2	251	23	ABP47870	Human polypeptide
35	7	2.2	298	23	ABP30505	Streptococcus poly
36	7	2.2	300	21	AB808497	Envelope 2 protein
37	7	2.2	304	17	AA898460	Endonuclease Ilati
38	7	2.2	304	17	AAW02163	Lactococcus lactis
39	7	2.2	321	23	ABP28875	Streptococcus poly
40	7	2.2	337	21	AA638356	Arabidopsis thalia
41	7	2.2	344	19	AA638355	Arabidopsis thalia
42	7	2.2	344	19	AAW69752	Acetobacter xyliu
43	7	2.2	355	19	AAV10975	H. pylori ORF hp5p
44	7	2.2	356	22	AA883221	Yeast FARP signatu
45	7	2.2	357	21	AA609523	Arabidopsis thalia
46	7	2.2	361	22	AB868291	Drosophila melanog
47	7	2.2	363	21	AA609522	Arabidopsis thalia
48	7	2.2	387	21	AAV93670	Amino acid sequenc
49	7	2.2	387	21	AAV93673	Amino acid sequenc
50	7	2.2	387	21	AAV93674	Amino acid sequenc
51	7	2.2	387	21	AAV93675	Amino acid sequenc
52	7	2.2	387	22	AA846854	S. cerevisiae RPL3
53	7	2.2	392	22	AB869425	Drosophila melanog
54	7	2.2	395	22	AB869583	Drosophila melanog
55	7	2.2	395	22	ABG10252	Novel human diagno
56	7	2.2	396	22	AB860029	Drosophila melanog
57	7	2.2	400	23	ABP29950	Streptococcus poly
58	7	2.2	401	21	AA638354	Arabidopsis thalia
59	7	2.2	421	21	AA609521	Arabidopsis thalia
60	7	2.2	427	22	AAU51308	Proionbacterium
61	7	2.2	452	21	AA620779	Arabidopsis thalia
62	7	2.2	452	21	AA638868	Arabidopsis thalia
63	7	2.2	460	21	AA620778	Arabidopsis thalia
64	7	2.2	460	21	AA638867	Arabidopsis thalia
65	7	2.2	468	21	AA620777	Arabidopsis thalia
66	7	2.2	468	21	AA638866	Arabidopsis thalia
67	7	2.2	499	23	AB847977	Listeria monocytog
68	7	2.2	501	17	AAW03760	Wallerian inhibiti
69	7	2.2	505	15	AA841921	MISR2A/MISR2B. Ra
70	7	2.2	505	15	AA855369	Human Activin rece
71	7	2.2	505	15	AA855373	Mouse Activin rece
72	7	2.2	505	16	AA870240	Serine/threonine k
73	7	2.2	505	20	AAV33302	Human hMLK-4 clone
74	7	2.2	505	20	AAV33306	Human hMLK-4 clone
75	7	2.2	532	22	AB864136	Drosophila melanog
76	7	2.2	532	22	AB866951	Drosophila melanog
77	7	2.2	543	18	AAW07702	Mouse FRS2 repress
78	7	2.2	571	22	AB862679	Drosophila melanog
79	7	2.2	623	20	AAV14967	Amino acid sequenc
80	7	2.2	623	22	AA883267	Yeast FARP SEQ ID
81	7	2.2	623	22	AA883275	Yeast FARP SEQ ID
82	7	2.2	674	22	ABG12515	Novel human diagno
83	7	2.2	687	21	AAV69165	A human N-acetylgl

84	7	2.2	727	22	ABG25530	Novel human diagno	157	6	1.9	14	23	AAU76046	Synthetic linker m
85	7	2.2	732	22	ABG10453	Novel human diagno	158	6	1.9	15	16	AAU79629	Endocarditis speci
86	7	2.2	753	23	ABP28191	Streptococcus poly	159	6	1.9	16	16	AAU79629	Endocarditis speci
87	7	2.2	761	23	ABP27624	Streptococcus poly	160	6	1.9	16	23	ABG62289	Endocarditis speci
88	7	2.2	765	21	AAV69167	A mature human N-a	161	6	1.9	16	23	ABG78470	Gum arabic glycopr
89	7	2.2	765	21	AAV69168	A mature human N-a	162	6	1.9	16	23	ABG78471	Gum arabic glycopr
90	7	2.2	775	22	AAU48754	Propionibacterium	163	6	1.9	16	23	AAU76047	Gum arabic glycopr
91	7	2.2	806	22	AAU48754	C glutamicum prote	164	6	1.9	16	23	AAU76048	Synthetic linker m
92	7	2.2	814	23	ABG92136	Herbicidially activ	165	6	1.9	18	14	AAU76368	Synthetic linker m
93	7	2.2	836	22	ABG67517	Drosophila melanog	166	6	1.9	18	15	AAU55861	Linker peptide. S
94	7	2.2	842	20	AAV66631	Drosophila melanog	167	6	1.9	18	15	AAU55862	Linker peptide. S
95	7	2.2	843	20	AAV66631	Chicken capsacin	168	6	1.9	18	20	AAU68785	Antibody-beta-lact
96	7	2.2	843	20	AAV66631	Chicken VRI capsai	169	6	1.9	18	20	AAU68785	Antibody-beta-lact
97	7	2.2	852	22	ABG64019	Drosophila melanog	170	6	1.9	18	20	AAU68785	Antibody-beta-lact
98	7	2.2	864	22	ABG71319	Drosophila melanog	171	6	1.9	18	20	AAU68785	Antibody-beta-lact
99	7	2.2	922	21	AAU68785	Candida albicans C	172	6	1.9	18	20	AAU68785	Antibody-beta-lact
100	7	2.2	955	22	ABG71635	Drosophila melanog	173	6	1.9	18	21	AAU68785	Antibody-beta-lact
101	7	2.2	960	23	AAU75476	S. aureus antigeni	174	6	1.9	18	21	AAU68785	Antibody-beta-lact
102	7	2.2	1047	22	ABG59191	Drosophila melanog	175	6	1.9	18	21	AAU68785	Antibody-beta-lact
103	7	2.2	1062	22	ABG71830	Drosophila melanog	176	6	1.9	18	21	AAU68785	Antibody-beta-lact
104	7	2.2	1336	23	AAU92973	Arabidopsis thalian	177	6	1.9	18	21	AAU68785	Antibody-beta-lact
105	7	2.2	1485	21	AAU69166	A mature human N-a	178	6	1.9	18	21	AAU68785	Antibody-beta-lact
106	7	2.2	2071	22	AAU69166	Mouse maeel protei	179	6	1.9	18	21	AAU68785	Antibody-beta-lact
107	7	2.2	2344	22	AAU69166	Mouse maeel protei	180	6	1.9	18	21	AAU68785	Antibody-beta-lact
108	7	2.2	2780	22	AAU69166	Mouse maeel protei	181	6	1.9	18	21	AAU68785	Antibody-beta-lact
109	7	2.2	2870	21	AAU69166	Mouse maeel protei	182	6	1.9	18	21	AAU68785	Antibody-beta-lact
110	7	2.2	3010	21	AAU69166	Mouse maeel protei	183	6	1.9	18	21	AAU68785	Antibody-beta-lact
111	7	2.2	3070	21	AAU69166	Mouse maeel protei	184	6	1.9	18	21	AAU68785	Antibody-beta-lact
112	7	2.2	3178	21	AAU69166	Mouse maeel protei	185	6	1.9	18	21	AAU68785	Antibody-beta-lact
113	7	2.2	3178	21	AAU69166	Mouse maeel protei	186	6	1.9	18	21	AAU68785	Antibody-beta-lact
114	7	2.2	3178	21	AAU69166	Mouse maeel protei	187	6	1.9	18	21	AAU68785	Antibody-beta-lact
115	7	2.2	3178	21	AAU69166	Mouse maeel protei	188	6	1.9	18	21	AAU68785	Antibody-beta-lact
116	7	2.2	3178	21	AAU69166	Mouse maeel protei	189	6	1.9	18	21	AAU68785	Antibody-beta-lact
117	7	2.2	3178	21	AAU69166	Mouse maeel protei	190	6	1.9	18	21	AAU68785	Antibody-beta-lact
118	7	2.2	3178	21	AAU69166	Mouse maeel protei	191	6	1.9	18	21	AAU68785	Antibody-beta-lact
119	7	2.2	3178	21	AAU69166	Mouse maeel protei	192	6	1.9	18	21	AAU68785	Antibody-beta-lact
120	7	2.2	3178	21	AAU69166	Mouse maeel protei	193	6	1.9	18	21	AAU68785	Antibody-beta-lact
121	7	2.2	3178	21	AAU69166	Mouse maeel protei	194	6	1.9	18	21	AAU68785	Antibody-beta-lact
122	7	2.2	3178	21	AAU69166	Mouse maeel protei	195	6	1.9	18	21	AAU68785	Antibody-beta-lact
123	7	2.2	3178	21	AAU69166	Mouse maeel protei	196	6	1.9	18	21	AAU68785	Antibody-beta-lact
124	7	2.2	3178	21	AAU69166	Mouse maeel protei	197	6	1.9	18	21	AAU68785	Antibody-beta-lact
125	7	2.2	3178	21	AAU69166	Mouse maeel protei	198	6	1.9	18	21	AAU68785	Antibody-beta-lact
126	7	2.2	3178	21	AAU69166	Mouse maeel protei	199	6	1.9	18	21	AAU68785	Antibody-beta-lact
127	7	2.2	3178	21	AAU69166	Mouse maeel protei	200	6	1.9	18	21	AAU68785	Antibody-beta-lact
128	7	2.2	3178	21	AAU69166	Mouse maeel protei	201	6	1.9	18	21	AAU68785	Antibody-beta-lact
129	7	2.2	3178	21	AAU69166	Mouse maeel protei	202	6	1.9	18	21	AAU68785	Antibody-beta-lact
130	7	2.2	3178	21	AAU69166	Mouse maeel protei	203	6	1.9	18	21	AAU68785	Antibody-beta-lact
131	7	2.2	3178	21	AAU69166	Mouse maeel protei	204	6	1.9	18	21	AAU68785	Antibody-beta-lact
132	7	2.2	3178	21	AAU69166	Mouse maeel protei	205	6	1.9	18	21	AAU68785	Antibody-beta-lact
133	7	2.2	3178	21	AAU69166	Mouse maeel protei	206	6	1.9	18	21	AAU68785	Antibody-beta-lact
134	7	2.2	3178	21	AAU69166	Mouse maeel protei	207	6	1.9	18	21	AAU68785	Antibody-beta-lact
135	7	2.2	3178	21	AAU69166	Mouse maeel protei	208	6	1.9	18	21	AAU68785	Antibody-beta-lact
136	7	2.2	3178	21	AAU69166	Mouse maeel protei	209	6	1.9	18	21	AAU68785	Antibody-beta-lact
137	7	2.2	3178	21	AAU69166	Mouse maeel protei	210	6	1.9	18	21	AAU68785	Antibody-beta-lact
138	7	2.2	3178	21	AAU69166	Mouse maeel protei	211	6	1.9	18	21	AAU68785	Antibody-beta-lact
139	7	2.2	3178	21	AAU69166	Mouse maeel protei	212	6	1.9	18	21	AAU68785	Antibody-beta-lact
140	7	2.2	3178	21	AAU69166	Mouse maeel protei	213	6	1.9	18	21	AAU68785	Antibody-beta-lact
141	7	2.2	3178	21	AAU69166	Mouse maeel protei	214	6	1.9	18	21	AAU68785	Antibody-beta-lact
142	7	2.2	3178	21	AAU69166	Mouse maeel protei	215	6	1.9	18	21	AAU68785	Antibody-beta-lact
143	7	2.2	3178	21	AAU69166	Mouse maeel protei	216	6	1.9	18	21	AAU68785	Antibody-beta-lact
144	7	2.2	3178	21	AAU69166	Mouse maeel protei	217	6	1.9	18	21	AAU68785	Antibody-beta-lact
145	7	2.2	3178	21	AAU69166	Mouse maeel protei	218	6	1.9	18	21	AAU68785	Antibody-beta-lact
146	7	2.2	3178	21	AAU69166	Mouse maeel protei	219	6	1.9	18	21	AAU68785	Antibody-beta-lact
147	7	2.2	3178	21	AAU69166	Mouse maeel protei	220	6	1.9	18	21	AAU68785	Antibody-beta-lact
148	7	2.2	3178	21	AAU69166	Mouse maeel protei	221	6	1.9	18	21	AAU68785	Antibody-beta-lact
149	7	2.2	3178	21	AAU69166	Mouse maeel protei	222	6	1.9	18	21	AAU68785	Antibody-beta-lact
150	7	2.2	3178	21	AAU69166	Mouse maeel protei	223	6	1.9	18	21	AAU68785	Antibody-beta-lact
151	7	2.2	3178	21	AAU69166	Mouse maeel protei	224	6	1.9	18	21	AAU68785	Antibody-beta-lact
152	7	2.2	3178	21	AAU69166	Mouse maeel protei	225	6	1.9	18	21	AAU68785	Antibody-beta-lact
153	7	2.2	3178	21	AAU69166	Mouse maeel protei	226	6	1.9	18	21	AAU68785	Antibody-beta-lact
154	7	2.2	3178	21	AAU69166	Mouse maeel protei	227	6	1.9	18	21	AAU68785	Antibody-beta-lact
155	7	2.2	3178	21	AAU69166	Mouse maeel protei	228	6	1.9	18	21	AAU68785	Antibody-beta-lact
156	7	2.2	3178	21	AAU69166	Mouse maeel protei	229	6	1.9	18	21	AAU68785	Antibody-beta-lact

230	6	1.9	23	23	ABR78497	Gum arabic glycop	303	6	1.9	67	21	AA02271	Human secreted pro
231	6	1.9	24	20	AA01271	Gum arabic glycop	304	6	1.9	68	22	AAU48057	Proionbacterium
232	6	1.9	24	20	ABR78498	Gum arabic glycop	305	6	1.9	69	22	AAU50589	Proionbacterium
233	6	1.9	32	21	AA095780	Glycosylphosphidyl	306	6	1.9	69	22	AA01887	Human bone marrow
234	6	1.9	32	22	ABR23454	Protein #5453 enco	307	6	1.9	69	23	ABG41700	Human peptide enco
235	6	1.9	32	22	AA058895	Human brain expres	308	6	1.9	70	22	AAU57491	Proionbacterium
236	6	1.9	32	22	AA01421	Human bone marrow	309	6	1.9	70	22	AAU60090	Proionbacterium
237	6	1.9	32	22	AA019067	Human bone marrow	310	6	1.9	70	22	AB010629	Human pancreatic c
238	6	1.9	32	23	ABG41225	Human peptide enco	311	6	1.9	70	22	AA067429	Human immune/haema
239	6	1.9	33	22	ABR38493	Protein #5599 enco	312	6	1.9	70	22	AA062385	Human digestive sy
240	6	1.9	33	22	ABR23645	Protein #5644 enco	313	6	1.9	71	22	AAU63040	Proionbacterium
241	6	1.9	33	22	AA059109	Human brain expres	314	6	1.9	72	22	AAU50948	Proionbacterium
242	6	1.9	33	22	AA017164	Human bone marrow	315	6	1.9	72	22	AAU66354	Proionbacterium
243	6	1.9	33	22	AA019248	Protein #5682 enco	316	6	1.9	73	22	AAU41987	Proionbacterium
244	6	1.9	33	22	AA019388	Protein #5975 enco	317	6	1.9	74	22	AAU63140	Proionbacterium
245	6	1.9	33	23	ABG41458	Human peptide enco	318	6	1.9	74	22	AA099770	Human excretory re
246	6	1.9	34	21	AA015436	TCR beta V-N-J reg	319	6	1.9	74	22	AA042585	Human kidney relat
247	6	1.9	34	22	AA013463	Human polypeptide	320	6	1.9	74	23	AA042585	Human kidney relat
248	6	1.9	34	22	AA077262	Human colon cancer	321	6	1.9	76	20	ABR78446	Gum arabic glycop
249	6	1.9	38	21	AA093764	Amino acid sequenc	322	6	1.9	76	22	AA098100	Yeast cdc24 matng
250	6	1.9	38	21	ABR78494	Gum arabic glycop	323	6	1.9	76	22	AAU54993	Proionbacterium
251	6	1.9	39	23	ABR78445	Gum arabic glycop	324	6	1.9	76	22	ABR43310	Protein #10816 enc
252	6	1.9	41	23	ABR81352	Human prostate spe	325	6	1.9	76	22	AA064218	Human brain expres
253	6	1.9	42	21	ABR53853	Human colon cancer	326	6	1.9	76	22	AA077043	Human bone marrow
254	6	1.9	43	20	AA014015	Yash repeat sequen	327	6	1.9	76	22	AA037157	Protein #11194 enc
255	6	1.9	44	20	AA013971	LysB repeat sequen	328	6	1.9	77	21	AA099883	Human ORFX protein
256	6	1.9	44	20	AA013973	XyIA repeat sequen	329	6	1.9	77	21	AA032894	Maize embryo spect
257	6	1.9	44	20	AA013969	Lys repeat sequenc	330	6	1.9	77	22	AA085089	Arabidopsis thalia
258	6	1.9	44	22	ABR50846	Human secreted pro	331	6	1.9	79	17	AA097976	Human immune/haema
259	6	1.9	45	23	ABR30650	Streptococcus poly	332	6	1.9	80	23	ABR07845	Fusion protein car
260	6	1.9	48	22	ABR42215	Protein #9721 enco	333	6	1.9	81	22	ABR57836	Human ORFX protein
261	6	1.9	48	22	AA063100	Human brain expres	334	6	1.9	82	22	AAU42827	Drosophila melanog
262	6	1.9	48	22	AA075911	Human bone marrow	335	6	1.9	82	23	ABR26938	Proionbacterium
263	6	1.9	48	22	AA036022	Protein #10059 enc	336	6	1.9	83	21	AA033795	Streptococcus poly
264	6	1.9	49	19	AA059911	Amino acid sequenc	337	6	1.9	83	22	AA066534	Arabidopsis thalia
265	6	1.9	50	22	AAU57285	Proionbacterium	338	6	1.9	83	22	AAU47318	Novel human connec
266	6	1.9	50	22	AAU63205	Proionbacterium	339	6	1.9	83	23	ABR31716	Proionbacterium
267	6	1.9	50	22	ABR35431	Human ORF4404 prot	340	6	1.9	84	22	AAU20521	Human ORF689 prote
268	6	1.9	51	22	AAU64877	Proionbacterium	341	6	1.9	85	22	ABG06252	Human secreted pro
269	6	1.9	52	17	AA020093	S. mutans antigen	342	6	1.9	85	23	ABR00054	Novel human diagno
270	6	1.9	53	20	AAU57802	Proionbacterium	343	6	1.9	86	21	AA06563	Human ORFX protein
271	6	1.9	54	20	AAU02724	Human secreted pro	344	6	1.9	86	21	AAU61512	C. elegans insulin
272	6	1.9	54	23	ABR34838	Human ORF3811 prot	345	6	1.9	87	22	ABR68252	Proionbacterium
273	6	1.9	55	22	ABR41095	Protein #8601 enco	346	6	1.9	87	23	ABR49133	Drosophila melanog
274	6	1.9	55	22	ABR25148	Protein #7147 enco	347	6	1.9	92	21	AA011138	Listeria monocytog
275	6	1.9	55	22	AA061951	Human brain expres	348	6	1.9	92	22	AAU51102	Human secreted pro
276	6	1.9	55	22	AA074755	Human bone marrow	349	6	1.9	92	22	AA097226	Proionbacterium
277	6	1.9	55	22	AA008095	Human polypeptide	350	6	1.9	93	22	AAU27941	Melanoma polypepto
278	6	1.9	55	22	AA020401	Protein #6835 enco	351	6	1.9	93	22	AAU09387	Human contig polyp
279	6	1.9	55	22	AA044871	Protein #8908 enco	352	6	1.9	93	22	AAU96807	Human polypeptide
280	6	1.9	55	23	ABG44552	Human peptide enco	353	6	1.9	94	22	AAU08047	Cancer gene therap
281	6	1.9	55	22	ABR05095	Human ORFX protein	354	6	1.9	95	21	AA057203	Human polypeptide
282	6	1.9	56	22	AAU55761	Proionbacterium	355	6	1.9	95	21	AA059156	Arabidopsis thalia
283	6	1.9	56	22	AA063969	Human immune/haema	356	6	1.9	96	22	AAU63242	Arabidopsis thalia
284	6	1.9	58	22	AAU67225	Proionbacterium	357	6	1.9	97	18	AAU44126	Proionbacterium
285	6	1.9	59	22	AAU54245	Proionbacterium	358	6	1.9	97	22	AAU47608	Streptococcus pneu
286	6	1.9	59	22	ABG33751	Novel human diagno	359	6	1.9	97	22	AAU66331	Proionbacterium
287	6	1.9	60	22	AAU39971	Proionbacterium	360	6	1.9	97	23	ABR47813	Proionbacterium
288	6	1.9	60	22	AA013850	Human polypeptide	361	6	1.9	98	21	AA015079	WntV protein relat
289	6	1.9	61	22	AA068541	Human immune/haema	362	6	1.9	98	22	AA083809	Arabidopsis thalia
290	6	1.9	61	23	ABR08454	Human ORFX protein	363	6	1.9	100	22	AAU59584	Human immune/haema
291	6	1.9	62	22	AAU52735	Proionbacterium	364	6	1.9	100	22	ABG59960	Proionbacterium
292	6	1.9	62	22	AAU01568	Human polypeptide	365	6	1.9	101	21	AA055982	Human DTRP polype
293	6	1.9	62	22	AA057773	Human colon cancer	366	6	1.9	101	22	AAU57377	Arabidopsis thalia
294	6	1.9	63	22	AAU39814	Proionbacterium	367	6	1.9	101	22	AAU92336	Proionbacterium
295	6	1.9	63	22	AAU51768	Proionbacterium	368	6	1.9	102	19	AA092336	C glutaminc prote
296	6	1.9	63	22	AA092833	C glutaminc prote	369	6	1.9	102	22	AA06407	S. pneumoniae deri
297	6	1.9	64	21	AA025878	Zea mays protein f	370	6	1.9	103	22	AA066912	Human foetal prote
298	6	1.9	64	22	AAU51537	Proionbacterium	371	6	1.9	103	22	AAU72557	Human immune/haema
299	6	1.9	64	22	AAU64736	Proionbacterium	372	6	1.9	104	21	AA080445	Escherichia coli s
300	6	1.9	65	20	AA088927	Sequence ID #627 f	373	6	1.9	104	21	AA034476	A human prostate s
301	6	1.9	65	20	ABR50956	Human secreted pro	374	6	1.9	104	21	AA045464	Arabidopsis thalia
302	6	1.9	66	22	AA062784	Human immune/haema	375	6	1.9	104	22	AAU42872	Proionbacterium

376	6	1.9	105	22	AAU50456	Protonibacterium	449	6	1.9	144	21	AA636306	Arabidopsis thalia
377	6	1.9	105	22	ABR29071	Peptide #1722 enco	450	6	1.9	144	21	AA68999	Hypervariable regi
378	6	1.9	105	22	ABR34231	Peptide #1739 enco	451	6	1.9	144	23	ABP18054	Staphylococcus epi
379	6	1.9	105	22	ABR50792	Human secreted pro	452	6	1.9	144	23	ABR48853	Listeria monocytog
380	6	1.9	105	22	AAW50027	Human brain exproe	453	6	1.9	145	21	AA633999	Arabidopsis thalia
381	6	1.9	105	22	AAW67412	Human bone marrow	454	6	1.9	145	23	ABP05983	Human ORFX protein
382	6	1.9	105	22	AAW65355	Human immune/haema	455	6	1.9	147	15	AAW60799	Rape abscission/de
383	6	1.9	105	22	AAW68807	Human immune/haema	456	6	1.9	148	21	AA619797	Arabidopsis thalia
384	6	1.9	105	22	AAW15243	Peptide #1577 enco	457	6	1.9	148	21	AA635884	Arabidopsis thalia
385	6	1.9	105	22	AAW27704	Peptide #1741 enco	458	6	1.9	148	21	AA637201	Arabidopsis thalia
386	6	1.9	106	22	AAW35541	Peptide #1668 enco	459	6	1.9	148	21	AA637201	Arabidopsis thalia
387	6	1.9	106	22	AAW43541	Human cancer assoc	460	6	1.9	148	21	AA638547	Arabidopsis thalia
388	6	1.9	106	22	AAW00089	Alpha-amylase sequ	461	6	1.9	148	21	AAW69006	Amino acid sequenc
389	6	1.9	106	22	AAW60827	Manduca sexta chlo	462	6	1.9	148	21	AAW69006	Novel human diagno
390	6	1.9	107	18	AAW16534	Human anti-RSV mon	463	6	1.9	150	21	AAW69726	Arabidopsis thalia
391	6	1.9	107	22	AAW52872	Escherichia coli p	464	6	1.9	150	22	AAW47724	Protonibacterium
392	6	1.9	107	22	AAW06855	Human foetal prote	465	6	1.9	150	22	AAW66351	Protonibacterium
393	6	1.9	108	22	AAW58598	Drosophila melanog	466	6	1.9	150	22	AAW66351	Human AFP protein
394	6	1.9	108	22	AAW68722	Drosophila melanog	467	6	1.9	151	21	AAW61368	Zea mays protein f
395	6	1.9	109	23	AAW06860	Human nCPER-Seg107	468	6	1.9	151	22	AAW69434	Drosophila melanog
396	6	1.9	110	21	AAW16254	Arabidopsis thalia	469	6	1.9	151	22	AAW48426	Protonibacterium
397	6	1.9	110	21	AAW23646	Arabidopsis thalia	470	6	1.9	151	22	AAW69434	C glutaminc prote
398	6	1.9	110	22	AAW60794	Protonibacterium	471	6	1.9	152	22	AAW626293	Novel human diagno
399	6	1.9	111	22	AAW49009	Protonibacterium	472	6	1.9	152	22	AAW626293	C glutaminc prote
400	6	1.9	111	23	AAW34224	Human ORF3197 prot	473	6	1.9	154	19	AAW723102	Novel human diagno
401	6	1.9	115	21	AAW33280	Plnus radiata tran	474	6	1.9	154	21	AAW62533	Arabidopsis thalia
402	6	1.9	115	23	AAW42590	Human ovarian anti	475	6	1.9	154	21	AAW62533	Arabidopsis thalia
403	6	1.9	116	16	AAW80187	MSP1-derived poly	476	6	1.9	155	22	AAW68650	Drosophila melanog
404	6	1.9	116	16	AAW80189	Human nervous syst	477	6	1.9	156	22	AAW91069	Streptomyces nogal
405	6	1.9	116	22	AAW92971	Human digestive sy	478	6	1.9	156	22	AAW91069	Novel human diagno
406	6	1.9	116	22	AAW92971	Human liver associ	479	6	1.9	159	21	AAW55981	Novel human diagno
407	6	1.9	116	22	AAW92971	Human liver associ	480	6	1.9	159	21	AAW55981	Human ORFX protein
408	6	1.9	116	22	AAW92971	Human liver associ	481	6	1.9	160	22	AAW34406	Protonibacterium
409	6	1.9	116	22	AAW92971	Human ORFX protein	482	6	1.9	161	21	AAW37770	Arabidopsis thalia
410	6	1.9	116	22	AAW92971	Human ORFX protein	483	6	1.9	161	21	AAW37770	Streptococcus pneu
411	6	1.9	117	22	AAW92971	Novel human diagno	484	6	1.9	161	22	AAW01086	Streptococcus pneu
412	6	1.9	117	22	AAW92971	Novel human diagno	485	6	1.9	162	22	AAW01086	CER 89 protein seq
413	6	1.9	117	22	AAW92971	Human secreted pro	486	6	1.9	162	22	AAW01086	Novel human diagno
414	6	1.9	119	22	AAW08893	Protonibacterium	487	6	1.9	162	22	AAW08893	Novel human diagno
415	6	1.9	120	22	AAW48739	Human ovarian anti	488	6	1.9	164	21	AAW37955	Arabidopsis thalia
416	6	1.9	120	23	AAW51633	Arabidopsis thalia	489	6	1.9	164	21	AAW37955	Arabidopsis thalia
417	6	1.9	121	21	AAW41143	Protonibacterium	490	6	1.9	166	22	AAW11597	Novel human diagno
418	6	1.9	121	22	AAW34475	Human protein sequ	491	6	1.9	166	22	AAW11597	Novel human diagno
419	6	1.9	122	22	AAW55052	Human prostate can	492	6	1.9	168	23	AAW87029	Adenovirus 5 fiber
420	6	1.9	123	21	AAW57118	H. pylori ORF 29ge	493	6	1.9	169	21	AAW81561	Human N-acetylgluc
421	6	1.9	124	18	AAW55414	Human pancreatic c	494	6	1.9	171	21	AAW81561	Eucalyptus grandis
422	6	1.9	124	21	AAW54240	Human radiata tran	495	6	1.9	172	22	AAW32645	Peptide #6528 enco
423	6	1.9	124	21	AAW53267	Human ORF662 prote	496	6	1.9	172	22	AAW32645	Human bone marrow
424	6	1.9	124	23	AAW31689	Novel human diagno	497	6	1.9	172	22	AAW32645	Human brain expres
425	6	1.9	125	22	AAW12237	Human polypeptide	498	6	1.9	172	22	AAW12237	Peptide #5947 enco
426	6	1.9	125	22	AAW06825	Human secreted pro	499	6	1.9	172	22	AAW06825	Peptide #6548 enco
427	6	1.9	127	21	AAW02197	Mitochondrial acyl	500	6	1.9	172	22	AAW02197	Human polypeptide
428	6	1.9	129	22	AAW80916	Human ORFX protein	501	6	1.9	175	23	AAW80916	Arabidopsis thalia
429	6	1.9	129	22	AAW80916	Human ORFX protein	502	6	1.9	175	23	AAW80916	Rattus norvegicus
430	6	1.9	130	22	AAW79711	Arabidopsis thalia	503	6	1.9	176	23	AAW79711	Physcomitrella pat
431	6	1.9	131	21	AAW34233	Human contig poly	504	6	1.9	177	22	AAW34233	Corynebacterium gl
432	6	1.9	131	22	AAW79711	Lipid biochemesis	505	6	1.9	177	22	AAW79711	Protonibacterium
433	6	1.9	133	22	AAW80853	P. patens lipid met	506	6	1.9	181	22	AAW80853	Arabidopsis thalia
434	6	1.9	133	22	AAW83294	Human LFA-3 D2 reg	507	6	1.9	181	22	AAW83294	Arabidopsis thalia
435	6	1.9	134	13	AAW55683	Protonibacterium	508	6	1.9	182	21	AAW55683	Amino acid sequenc
436	6	1.9	134	13	AAW55683	Human polypeptide	509	6	1.9	182	21	AAW55683	Neisseria meningit
437	6	1.9	134	23	AAW88916	Human LFA-3-delta	510	6	1.9	183	21	AAW88916	Human polypeptide
438	6	1.9	135	13	AAW28369	Human LFA-3-delta	511	6	1.9	185	21	AAW28369	Drosophila melanog
439	6	1.9	135	17	AAW94881	Actinomadura flexu	512	6	1.9	185	21	AAW94881	Novel human diagno
440	6	1.9	135	22	AAW63617	Protonibacterium	513	6	1.9	187	22	AAW63617	Herbicidally activ
441	6	1.9	137	22	AAW24903	Novel human diagno	514	6	1.9	187	22	AAW24903	Novel human diagno
442	6	1.9	137	22	AAW06625	Human ORFX protein	515	6	1.9	187	22	AAW06625	Novel human diagno
443	6	1.9	140	21	AAW33784	Arabidopsis thalia	516	6	1.9	189	22	AAW33784	Novel human diagno
444	6	1.9	140	21	AAW33784	Arabidopsis thalia	517	6	1.9	191	22	AAW33784	Novel human diagno
445	6	1.9	142	22	AAW65722	Drosophila melanog	518	6	1.9	193	23	AAW65722	Human polypeptide
446	6	1.9	142	22	AAW71893	C. glutaminc meta	519	6	1.9	193	23	AAW71893	Arabidopsis thalia
447	6	1.9	142	22	AAW79665	Corynebacterium gl	520	6	1.9	194	21	AAW79665	Amino acid sequenc
448	6	1.9	142	23	AAW53493	Lactococcus lactis	521	6	1.9	194	22	AAW53493	Arabidopsis thalia

522	6	1.9	196	22	AU20445
523	6	1.9	198	22	AAE02175
524	6	1.9	198	22	AB78986
525	6	1.9	199	22	ABG13674
526	6	1.9	199	22	AAE02180
527	6	1.9	199	22	AAE02182
528	6	1.9	200	22	ABG07420
529	6	1.9	200	22	ABG07860
530	6	1.9	200	22	ABP29630
531	6	1.9	202	22	ABBS9228
532	6	1.9	202	22	AU57004
533	6	1.9	202	22	ABP03464
534	6	1.9	203	20	AA37646
535	6	1.9	203	22	AAAG0062
536	6	1.9	205	22	AU34710
537	6	1.9	206	21	AA319275
538	6	1.9	207	23	AA50648
539	6	1.9	208	22	AU22481
540	6	1.9	208	22	AAAG1792
541	6	1.9	209	20	AA37892
542	6	1.9	209	22	AU87471
543	6	1.9	209	22	AU19732
544	6	1.9	209	23	ABP47952
545	6	1.9	211	20	AA35676
546	6	1.9	211	20	AA35677
547	6	1.9	211	20	AA305952
548	6	1.9	214	20	AA329117
549	6	1.9	215	22	ABBS9408
550	6	1.9	216	22	AU50081
551	6	1.9	218	21	AA308446
552	6	1.9	218	21	AA308447
553	6	1.9	220	19	AA369714
554	6	1.9	220	22	ABBS1843
555	6	1.9	220	22	AA386123
556	6	1.9	221	22	AA366635
557	6	1.9	222	15	AA360313
558	6	1.9	222	15	AA360589
559	6	1.9	223	22	AA388334
560	6	1.9	224	21	AA380818
561	6	1.9	225	21	AA313478
562	6	1.9	227	20	AA388925
563	6	1.9	227	22	ABBS0994
564	6	1.9	228	22	AU33676
565	6	1.9	230	22	AU54317
566	6	1.9	231	18	AA355544
567	6	1.9	231	20	AA371770
568	6	1.9	231	23	ABBS3629
569	6	1.9	233	22	AU25694
570	6	1.9	234	18	AA311638
571	6	1.9	234	23	AB347971
572	6	1.9	236	15	AA345442
573	6	1.9	236	20	AA383968
574	6	1.9	236	21	AA34602
575	6	1.9	236	22	AA38307
576	6	1.9	236	22	AA388881
577	6	1.9	237	12	AA34182
578	6	1.9	237	13	AA30804
579	6	1.9	237	17	AA30804
580	6	1.9	237	17	AA34020
581	6	1.9	237	18	AA31687
582	6	1.9	237	18	AA31585
583	6	1.9	237	19	AA30441
584	6	1.9	237	20	AA36189
585	6	1.9	237	20	AA35440
586	6	1.9	237	21	AA36127
587	6	1.9	237	21	AA319274
588	6	1.9	237	21	AA379059
589	6	1.9	237	21	AA34836
590	6	1.9	237	22	AU24035
591	6	1.9	237	22	AU2436
592	6	1.9	238	15	AA34543
593	6	1.9	238	21	AA31516
594	6	1.9	238	22	ABBS6978

Human secreted pro
Mycobacterium xeno
Lolium perenne lrp4
Novel human diagn
Mycobacterium xeno
Mycobacterium xeno
Novel human diagno
Novel human diagno
Streptococcus poly
Drosophila melanog
Propionibacterium
Human ORF protein
Amino acid sequenc
C glutamincum prote
E. coli cellular p
Arabidopsis thalia
Arabidopsis thalia
Novel human secret
C glutamincum prote
Lactobacillus brev
Novel central nerv
Human novel extrac
Human polypeptide
Eurogllyphus sp. al
Eurogllyphus sp. al
Protein deduced fr
Amino acid sequenc
Drosophila melanog
Propionibacterium
A human prostate s
A human protein havi
E. maynei Group 1
Arabidopsis thalia
Arabidopsis thalia
Polyptide fragme
Human secreted pro
Pseudomonas aerugi
Propionibacterium
H. pylori ORF 06ce
H. pylori outer me
Lactococcus lactis
G protein-coupled
Human anti-RSV mon
Listeria monocytog
Sequence of the si
Plasamid pBBP21 Dsb
Amino acid sequenc
Escherichia coli p
E. coli growth and
Human LFA-3 (CD58)
Human LFA-3 antige
Human LFA-3 antige
Anti-erbB2 scFv.
Human CD58 GPI. H
Single-chain anti-
Human LFA-3 antige
Human LFA-3 antige
A33/212 single-cha
Human cell surface
Arabidopsis thalia
Amino acid sequenc
Linked fusion prot
Epstein-Barr virus
Human lymphocyte f
Sequence of the si
Arabidopsis thalia
Drosophila melanog

595	6	1.9	238	23	AU92966
596	6	1.9	239	18	AA31516
597	6	1.9	240	11	AA305572
598	6	1.9	240	11	AA307604
599	6	1.9	240	11	AA327162
600	6	1.9	240	13	AA34222
601	6	1.9	240	14	AA34372
602	6	1.9	240	20	AA388035
603	6	1.9	240	21	AA37675
604	6	1.9	240	21	AA307931
605	6	1.9	240	21	AA313134
606	6	1.9	240	21	AA37250
607	6	1.9	240	21	AA380920
608	6	1.9	240	22	AA361158
609	6	1.9	240	23	AU76226
610	6	1.9	241	14	AA37645
611	6	1.9	241	20	AA35441
612	6	1.9	241	21	AA317852
613	6	1.9	241	21	AA354837
614	6	1.9	242	21	AA337570
615	6	1.9	242	22	ABBS3185
616	6	1.9	242	22	ABBS0561
617	6	1.9	243	20	AA329531
618	6	1.9	243	22	ABBS0343
619	6	1.9	243	22	AA390979
620	6	1.9	244	20	AA388096
621	6	1.9	244	21	AA327676
622	6	1.9	244	21	AA307932
623	6	1.9	244	21	AA37251
624	6	1.9	244	21	AA380921
625	6	1.9	244	21	AA352290
626	6	1.9	245	20	AA37891
627	6	1.9	247	21	AA313477
628	6	1.9	248	14	AA337646
629	6	1.9	249	20	AA37886
630	6	1.9	249	20	AA37886
631	6	1.9	250	9	AA31507
632	6	1.9	250	13	AA327161
633	6	1.9	250	14	AA34221
634	6	1.9	250	14	AA34371
635	6	1.9	250	14	AA37647
636	6	1.9	250	15	AA364271
637	6	1.9	250	17	AA304370
638	6	1.9	250	20	AA388097
639	6	1.9	250	21	AA327677
640	6	1.9	250	21	AA307933
641	6	1.9	250	21	AA317851
642	6	1.9	250	21	AA380817
643	6	1.9	250	21	AA383133
644	6	1.9	250	21	AA37252
645	6	1.9	250	21	AA380922
646	6	1.9	250	22	AA367519
647	6	1.9	250	22	AA361157
648	6	1.9	250	23	AA393167
649	6	1.9	250	23	AA376225
650	6	1.9	251	16	AA380192
651	6	1.9	251	18	AA302890
652	6	1.9	251	20	AA374015
653	6	1.9	251	22	ABBS03691
654	6	1.9	251	22	AA30451
655	6	1.9	251	22	AA320361
656	6	1.9	251	23	ABBS0062
657	6	1.9	252	20	AA30023
658	6	1.9	252	22	ABBS9632
659	6	1.9	252	22	AA36321
660	6	1.9	252	23	ABP43442
661	6	1.9	253	20	AA388098
662	6	1.9	253	21	AA327678
663	6	1.9	253	21	AA307934
664	6	1.9	253	21	AA303712
665	6	1.9	253	21	AA37253
666	6	1.9	253	21	AA380923
667	6	1.9	254	21	AA319874

Arabidopsis transc
Single-chain anti-
PI-linked LFA-3 en
Lymphocyte functio
PI-linked LFA-3 us
Amino acid sequenc
Sequence encoded b
Single-chain prote
Bivalent antigen b
A heterobivalent a
PI-linked human tr
4-4-20 VI region a
Single chain antig
Human PI-linked tr
PI-linked lymphocy
Sequence of a mult
A33/218 single-cha
Arabidopsis thalia
Linked fusion prot
Arabidopsis thalia
Drosophila melanog
Drosophila melanog
Human MEGP1 protei
Drosophila melanog
C glutamincum prote
Single-chain prote
Bivalent antigen b
A heterobivalent a
4-4-20 Vh region a
Single chain antig
Hamster acrosomal
A33/218 single cha
Sequence of a sing
CC49/218 single ch
Bacillus cereus ox
Human LFA-3 expres
Transmembrane LFA-
Amino acid sequenc
Sequence encoded b
Sequence of a 4-4-
Human LFA-3. Homo
Human cell adhesio
Single chain prote
4-4-20/212 single
A single-chain ant
Arabidopsis thalia
Arabidopsis thalia
Human transmembran
4-4-20/212 single-
Single chain antig
Amino acid sequenc
Human transmembran
Arabidopsis transmem
Lymphocyte functio
Minor ampullate sp
H. pylori cytoplasm
Human prostate tum
Novel human diagno
Human BP41608HDF
Human hepatoma-der
Streptococcus poly
Enterococcus faeca
Drosophila melanog
Putative P. abyssi
E faecalis BP110 a
Single chain prote
4-4-20/212 single
A single-chain ant
Arabidopsis thalia
4-4-20/212 single-
Single chain antig
Arabidopsis thalia

668	1.9	254	21	AAV69002	Amino acid sequenc	741	395	14	AAE41347	Nastertium xyloglu
669	1.9	254	23	ABP00351	Human ORFX protein	742	295	22	AAE4951	Shrimp white spot
670	1.9	255	21	AAE1515	Arabidopsis thalia	743	296	21	AAE13289	Arabidopsis thalia
671	1.9	255	23	AAU2763	Tomato DMT protein	744	296	21	AAE13289	Arabidopsis thalia
672	1.9	256	22	AAE2766	C glutamatum prote	745	296	21	AAE13289	Arabidopsis thalia
673	1.9	256	23	ABE2092	Herbicideally activ	746	296	21	AAE13289	Arabidopsis thalia
674	1.9	256	23	ABE2092	Herbicideally activ	747	297	21	AAE13289	Arabidopsis thalia
675	1.9	257	20	AAE1524	Single chain-antic	748	297	21	AAE13289	Arabidopsis thalia
676	1.9	257	20	AAE1524	Single chain-antic	749	297	21	AAE13289	Arabidopsis thalia
677	1.9	257	21	AAE1514	Arabidopsis thalia	750	297	21	AAE13289	Arabidopsis thalia
678	1.9	257	21	AAE1514	Arabidopsis thalia	751	297	21	AAE13289	Arabidopsis thalia
679	1.9	257	22	AAE2902	C glutamatum prote	752	298	22	AAE13289	Arabidopsis thalia
680	1.9	257	22	AAE2902	C glutamatum prote	753	299	21	AAE13289	Arabidopsis thalia
681	1.9	258	23	AAU10387	Trichoderma harzia	754	300	21	AAE13289	Arabidopsis thalia
682	1.9	258	23	AAU10387	Trichoderma harzia	755	301	21	AAE13289	Arabidopsis thalia
683	1.9	258	23	AAU10387	Trichoderma harzia	756	301	21	AAE13289	Arabidopsis thalia
684	1.9	259	22	AAE47418	Epse. Lactococcus	757	303	23	AAE13289	Arabidopsis thalia
685	1.9	259	22	AAE47418	Epse. Lactococcus	758	303	23	AAE13289	Arabidopsis thalia
686	1.9	262	18	AAE20758	H. pylori cytoplas	759	306	22	AAE13289	Arabidopsis thalia
687	1.9	262	20	AAE20758	H. pylori cytoplas	760	306	22	AAE13289	Arabidopsis thalia
688	1.9	262	21	AAE20758	H. pylori cytoplas	761	308	21	AAE13289	Arabidopsis thalia
689	1.9	262	21	AAE20758	H. pylori cytoplas	762	310	23	AAE13289	Arabidopsis thalia
690	1.9	263	12	AAE15055	Om212 single chain	763	312	17	AAE13289	Arabidopsis thalia
691	1.9	263	12	AAE15055	Om212 single chain	764	312	17	AAE13289	Arabidopsis thalia
692	1.9	264	17	AAE97381	Kabat consensus VK	765	312	23	AAE13289	Arabidopsis thalia
693	1.9	264	17	AAE97381	Kabat consensus VK	766	313	23	AAE13289	Arabidopsis thalia
694	1.9	264	19	AAE73049	Humantised A33 scab	767	314	16	AAE13289	Arabidopsis thalia
695	1.9	265	23	AAU75159	Novel floral meris	768	314	16	AAE13289	Arabidopsis thalia
696	1.9	265	23	AAU75159	Novel floral meris	769	314	16	AAE13289	Arabidopsis thalia
697	1.9	267	23	AAE43823	Arabidopsis thalia	770	314	16	AAE13289	Arabidopsis thalia
698	1.9	267	23	AAE43823	Arabidopsis thalia	771	314	16	AAE13289	Arabidopsis thalia
699	1.9	268	19	AAE60119	Human DTHP polype	772	314	16	AAE13289	Arabidopsis thalia
700	1.9	268	19	AAE60119	S. pneumoniae dexi	773	314	16	AAE13289	Arabidopsis thalia
701	1.9	268	22	AAE60931	Drosophila melanog	774	315	22	AAE13289	Arabidopsis thalia
702	1.9	268	22	AAE60931	Novel floral meris	775	315	22	AAE13289	Arabidopsis thalia
703	1.9	268	23	AAE60931	Novel floral meris	776	316	22	AAE13289	Arabidopsis thalia
704	1.9	269	23	AAE60931	Novel floral meris	777	317	22	AAE13289	Arabidopsis thalia
705	1.9	270	21	AAE19818	Adenovirus 5 recom	778	318	20	AAE13289	Arabidopsis thalia
706	1.9	271	21	AAE19818	Adenovirus 5 recom	779	319	22	AAE13289	Arabidopsis thalia
707	1.9	271	21	AAE19818	Adenovirus 5 recom	780	319	22	AAE13289	Arabidopsis thalia
708	1.9	271	21	AAE19818	Adenovirus 5 recom	781	320	21	AAE13289	Arabidopsis thalia
709	1.9	271	21	AAE19818	Adenovirus 5 recom	782	320	21	AAE13289	Arabidopsis thalia
710	1.9	271	21	AAE19818	Adenovirus 5 recom	783	321	22	AAE13289	Arabidopsis thalia
711	1.9	272	22	AAE19818	Adenovirus 5 recom	784	321	22	AAE13289	Arabidopsis thalia
712	1.9	274	21	AAE19818	Adenovirus 5 recom	785	321	22	AAE13289	Arabidopsis thalia
713	1.9	275	21	AAE19818	Adenovirus 5 recom	786	321	22	AAE13289	Arabidopsis thalia
714	1.9	275	21	AAE19818	Adenovirus 5 recom	787	321	22	AAE13289	Arabidopsis thalia
715	1.9	276	22	AAE19818	Adenovirus 5 recom	788	321	22	AAE13289	Arabidopsis thalia
716	1.9	276	22	AAE19818	Adenovirus 5 recom	789	322	21	AAE13289	Arabidopsis thalia
717	1.9	279	20	AAE19818	Adenovirus 5 recom	790	322	21	AAE13289	Arabidopsis thalia
718	1.9	279	20	AAE19818	Adenovirus 5 recom	791	322	21	AAE13289	Arabidopsis thalia
719	1.9	279	21	AAE19818	Adenovirus 5 recom	792	323	22	AAE13289	Arabidopsis thalia
720	1.9	279	21	AAE19818	Adenovirus 5 recom	793	324	23	AAE13289	Arabidopsis thalia
721	1.9	279	21	AAE19818	Adenovirus 5 recom	794	324	23	AAE13289	Arabidopsis thalia
722	1.9	279	21	AAE19818	Adenovirus 5 recom	795	325	23	AAE13289	Arabidopsis thalia
723	1.9	280	22	AAE19818	Adenovirus 5 recom	796	325	23	AAE13289	Arabidopsis thalia
724	1.9	281	23	AAE19818	Adenovirus 5 recom	797	326	21	AAE13289	Arabidopsis thalia
725	1.9	282	23	AAE19818	Adenovirus 5 recom	798	326	21	AAE13289	Arabidopsis thalia
726	1.9	283	20	AAE19818	Adenovirus 5 recom	799	327	22	AAE13289	Arabidopsis thalia
727	1.9	283	20	AAE19818	Adenovirus 5 recom	800	327	22	AAE13289	Arabidopsis thalia
728	1.9	284	20	AAE19818	Adenovirus 5 recom	801	328	22	AAE13289	Arabidopsis thalia
729	1.9	284	20	AAE19818	Adenovirus 5 recom	802	328	22	AAE13289	Arabidopsis thalia
730	1.9	285	22	AAE19818	Adenovirus 5 recom	803	329	22	AAE13289	Arabidopsis thalia
731	1.9	285	22	AAE19818	Adenovirus 5 recom	804	329	22	AAE13289	Arabidopsis thalia
732	1.9	286	22	AAE19818	Adenovirus 5 recom	805	329	22	AAE13289	Arabidopsis thalia
733	1.9	286	22	AAE19818	Adenovirus 5 recom	806	330	23	AAE13289	Arabidopsis thalia
734	1.9	287	20	AAE19818	Adenovirus 5 recom	807	330	23	AAE13289	Arabidopsis thalia
735	1.9	287	20	AAE19818	Adenovirus 5 recom	808	331	20	AAE13289	Arabidopsis thalia
736	1.9	290	21	AAE19818	Adenovirus 5 recom	809	331	20	AAE13289	Arabidopsis thalia
737	1.9	290	21	AAE19818	Adenovirus 5 recom	810	331	21	AAE13289	Arabidopsis thalia
738	1.9	292	22	AAE19818	Adenovirus 5 recom	811	331	22	AAE13289	Arabidopsis thalia
739	1.9	293	22	AAE19818	Adenovirus 5 recom	812	331	22	AAE13289	Arabidopsis thalia
740	1.9	295	11	AAE19818	Adenovirus 5 recom	813	331	23	AAE13289	Arabidopsis thalia

960	6	1.9	402	22	ABR64529	Drosophila melanog
961	6	1.9	403	21	ABG20948	Arabidopsis thaliana
962	6	1.9	403	21	AA629362	Arabidopsis thaliana
963	6	1.9	403	22	AA629362	C. glutamicum ATCC
964	6	1.9	405	22	ABG26543	Novel human diseno
965	6	1.9	407	21	AA772771	Streptomyces coelic
966	6	1.9	407	21	AA778823	Amino acid sequenc
967	6	1.9	407	22	AAE13679	C. glutamicum prote
968	6	1.9	407	23	AAE13583	Streptomyces fradi
969	6	1.9	408	23	ABB53521	Lactococcus lactis
970	6	1.9	409	22	AAE93991	Human protein sequ
971	6	1.9	410	21	AA733307	Rice acid triacylg
972	6	1.9	411	18	AAW01558	TYPI protein. Can
973	6	1.9	411	21	AAV15114	Candida cell-cycle
974	6	1.9	411	22	ABW71197	Drosophila melanog
975	6	1.9	412	22	AAE78973	C. glutamicum SRT
976	6	1.9	413	22	AAE00867	Human brain immuno
977	6	1.9	415	22	AAU17462	Novel signal trans
978	6	1.9	415	23	AAE66148	45 kDa protein of
979	6	1.9	416	19	AAW60836	Human LAMP family
980	6	1.9	416	20	AAE55927	Human STRK2 protei
981	6	1.9	416	20	AAE29605	Human TSC403 prote
982	6	1.9	416	20	AAE22648	A human signal tra
983	6	1.9	416	20	AAE22651	Marine signal tr
984	6	1.9	416	20	AAE21673	Human Ste20-like s
985	6	1.9	416	21	AAE82274	Mouse protein sequ
986	6	1.9	416	22	AAE82277	Human polypeptide
987	6	1.9	416	22	AAE82277	Human h2252 protei
988	6	1.9	416	22	AAE82277	Human TGR-beta rec
989	6	1.9	416	22	AAE82277	Corn putative leci
990	6	1.9	417	21	AAE82277	Human alternative
991	6	1.9	417	22	AAE82277	Human WTL variant
992	6	1.9	417	22	AAE82277	Comamonas testoste
993	6	1.9	418	15	AAE60135	Lipoleic-acid-dese
994	6	1.9	418	15	AAE60135	M. gondoniae gyrase
995	6	1.9	420	18	AAW15076	M. flavescentis gyra
996	6	1.9	420	18	AAW15075	M. flavescentis gyra
997	6	1.9	421	18	AAW15075	Drosophila melanog
998	6	1.9	421	21	AAE64654	Arabidopsis thaliana
999	6	1.9	422	21	AAE64654	Human polypeptide,
1000	6	1.9	422	22	AAE64654	

ALIGNMENTS

RESULT 1
ID AAB20106 standard; Protein; 322 AA.

AC AAB20106;

DT 23-APR-2001 (first entry)

DE Moraxella catarrhalis BASB110 protein.

KM BASB110: infection; otitis media; pneumonia; therapy; diagnosis;

KW antibacterial; antimicrobial; vaccine.

OS Moraxella catarrhalis.

FT Key Location/Qualifiers

FT Misc-difference 104 /note="Gly in translation of BASB110 PCR product"

XX WO200100838-A1.

XX 04-JAN-2001.

XX 23-JUN-2000; 2000WO-EP05854.

XX 25-JUN-1999; 99GB-0015031.

PA	(SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
XX	Thomard J;
XX	WPI: 2001-112459/12.
XX	N-PSDB; AAF30046.
XX	Novel BASB110 polypeptides of Moraxella catarrhalis, useful as a
XX	vaccine for treating Moraxella catarrhalis infections -
XX	Claim 1; Page 82-83; 88pp; English.
XX	The present sequence is that of BASB110 protein from Moraxella
XX	catarrhalis strain Mc2931 (ATCC 43617), a causative agent of otitis
XX	media in children and pneumonia in adults. The sequence is a
XX	translation of the BASB110 gene coding region (see AAF30046). It
XX	differs by 1 residue from the predicted polypeptide sequence (see
XX	AAE20107) of a BASB110 PCR product, having Ser rather than Gly at
XX	position 104. The invention provides BASB110 polypeptides, and
XX	polynucleotides encoding them, as well as expression vectors, host
XX	cells and methods for producing BASB110 polypeptides using
XX	recombinant methods. Also claimed is a vaccine composition
XX	comprising a BASB110 polypeptide, an immunogenic fragment of a
XX	BASB110 polypeptide, or a polypeptide having at least 85% amino
XX	acid sequence identity to BASB110, or comprising a polynucleotide
XX	encoding such a polypeptide. A claimed method of diagnosing a
XX	Moraxella infection involves identifying a BASB110 polypeptide or
XX	antibody. A claimed therapeutic composition useful in treating
XX	humans with M. catarrhalis infection comprises at least 1 antibody
XX	directed against a BASB110 polypeptide. BASB110 polypeptides also
XX	have utility in raising specific antibodies, and in screening for
XX	antibacterial drugs.
XX	Sequence 322 AA;
XX	Query Match 100.0%; Score 322; DB 22; Length 322;
XX	Best Local Similarity 100.0%; Pred. No. 1.2e-310; Gaps 0;
XX	Matches 322; Conservative 0; Mismatches 0; Indels 0;

QY 1 MVTITAINSONQKPKRIGLIFGVITTCITLACGSKPTYNSTSGSHRTSGSGIAGS 60

DB 1 MVTITAINSONQKPKRIGLIFGVITTCITLACGSKPTYNSTSGSHRTSGSGIAGS 60

QY 61 QVITDSQGVNRYOVKQDVTYSKIAORYGLNWRREIGHINNINSSTYITGMLTMSGDL 120

DB 61 QVITDSQGVNRYOVKQDVTYSKIAORYGLNWRREIGHINNINSSTYITGMLTMSGDL 120

QY 121 KYRERSISSGVNTAHTSPVAVOSSRPVQOHPAVOKETPPVYVYKKEPTPPVVOQAP 180

DB 121 KYRERSISSGVNTAHTSPVAVOSSRPVQOHPAVOKETPPVYVYKKEPTPPVVOQAP 180

QY 181 VAPPTTEAPFATGSGGVQFRYPVATVPVRRRGATVAGSTVYNSGMWFSGRGDILIN 240

DB 181 VAPPTTEAPFATGSGGVQFRYPVATVPVRRRGATVAGSTVYNSGMWFSGRGDILIN 240

QY 241 ASNAGTVIQADHNDGASIVIOHTNGFVSYTHIKDAOVKGTGVRTGQRTASMNQPSG 300

DB 241 ASNAGTVIQADHNDGASIVIOHTNGFVSYTHIKDAOVKGTGVRTGQRTASMNQPSG 300

QY 301 AALFEFRISRNQGVYVDPPLTVLK 322

DB 301 AALFEFRISRNQGVYVDPPLTVLK 322

RESULT 2

ID AAB20107 standard; Protein; 322 AA.

AC AAB20107;

DT 23-APR-2001 (first entry)

DE Moraxella catarrhalis BASB110 protein.

```

XX  BASB110; infection; otitis media; pneumonia; therapy; diagnosis;
KW  antibacterial; antimicrobial; vaccine.
XX
XX  Moraxella catarrhalis.
OS
XX
FH  Key Location/Qualifiers
FT  Misc-difference 104
FT  /note="Ser in translation of BASB110 gene"
XX
XX  WO200100838-A1.
XX
XX  04-JAN-2001.
XX
XX  23-JUN-2000; 2000WO-EP05854.
XX
XX  25-JUN-1999; 99GB-0015031.
XX
XX  (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
XX  Thonard J;
XX
XX  WPI: 2001-112459/12.
XX
XX  N-PSDB; AAF30047.
XX
XX  Novel BASB110 polypeptides of Moraxella catarrhalis, useful as a
XX  vaccine for treating Moraxella catarrhalis infections -
XX
XX  Claim 1; Page 83-84; 88pp; English.
XX
XX  The present sequence is that of BASB110 protein from Moraxella
XX  catarrhalis strain Mc231 (ATCC 43617), a causative agent of otitis
XX  media in children and pneumonia in adults. The sequence is a
XX  translation of a BASB110 PCR product (see AAF30047). It differs by
XX  1 residue from the predicted polypeptide sequence (see AAB20106) of
XX  the BASB110 gene product, having Gly rather than Ser at position
XX  104. The invention provides BASB110 polypeptides, and
XX  polynucleotides encoding them, as well as expression vectors, host
XX  cells and methods for producing BASB110 polypeptides using
XX  recombinant methods. Also claimed is a vaccine composition
XX  comprising a BASB110 polypeptide, an immunogenic fragment of a
XX  BASB110 polypeptide, or a polypeptide having at least 85% amino
XX  acid sequence identity to BASB110, or comprising a polynucleotide
XX  encoding such a polypeptide. A claimed method of diagnosing a
XX  Moraxella infection involves identifying a BASB110 polypeptide or
XX  antibody. A claimed therapeutic composition useful in treating
XX  humans with M. catarrhalis infection comprises at least 1 antibody
XX  directed against a BASB110 polypeptide. BASB110 polypeptides also
XX  have utility in raising specific antibodies, and in screening for
XX  antibacterial drugs.
XX
XX  Sequence 322 AA;
SQ
Query Match 68.6%; Score 221; DB 22; Length 322;
Best Local Similarity 99.7%; Pred. No. 1.7e-210;
Matches 321; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MTVTIATNSQOKPIKRLGIFGVITTCIAGCASKPTYNSTSGSGSHRTSGSGLAIGS 60
DB 1 MTVTIATNSQOKPIKRLGIFGVITTCIAGCASKPTYNSTSGSGSHRTSGSGLAIGS 60
QY 61 QVITDSQGVNRPVYKQGDVSKIAQRXGLNMRREIGHINNNSYTIYTGOMLTMSGD 120
DB 61 QVITDSQGVNRPVYKQGDVSKIAQRXGLNMRREIGHINNNSYTIYTGOMLTMSGD 120
QY 121 KVRERSISISSGVNATHTSPVAVQSSRPVQCHPAVQKPTPPVVVKKPTPPPVQCPAP 180
DB 121 KVRERSISISSGVNATHTSPVAVQSSRPVQCHPAVQKPTPPVVVKKPTPPPVQCPAP 180
QY 181 VAPVTEAPFATGSGVWQGFYPVGAATNPVVRRTAATVASTYNSNGMFFSGRDGLIN 240
DB 181 VAPVTEAPFATGSGVWQGFYPVGAATNPVVRRTAATVASTYNSNGMFFSGRDGLIN 240

```

```

QY 241 ASNAGTIVQADHNDGASIVIQHTNGFVSSYIHIKDAQVKTGDTVRTGQRIASMKNOPSG 300
DB 241 ASNAGTIVQADHNDGASIVIQHTNGFVSSYIHIKDAQVKTGDTVRTGQRIASMKNOPSG 300
QY 301 AALPEFRISRNGVYVDPPLTVLK 322
DB 301 AALPEFRISRNGVYVDPPLTVLK 322
RESULT 3
ABB61691
ID ABB61691 standard; Protein; 968 AA.
XX
XX  ABB61691;
AC
XX  26-MAR-2002 (first entry)
DT
XX
XX  Drosophila melanogaster polypeptide SEQ ID NO 11865.
DE
XX  Drosophila; developmental biology; cell signalling; insecticide;
KW  pharmaceutical.
XX
XX  Drosophila melanogaster.
OS
XX  WO200171042-A2.
XX
XX  23-MAR-2001; 2001WO-US09231.
XX
XX  23-MAR-2000; 2000US-191637P.
XX  11-JUL-2000; 2000US-0614150.
XX
XX  (PEKE ) PE CORP NY.
XX
XX  Venter JC, Adams M, Li PWD, Myers EW;
XX  WPI: 2001-656860/75.
XX  N-PSDB; ABL05794.
DR
XX
XX  New isolated nucleic acid detection reagent for detecting 1000 or more
XX  genes from Drosophila and for elucidating cell signalling and cell-cell
XX  interactions -
XX
XX  Disclosure; SEQ ID NO 11865; 21pp + Sequence Listing; English.
XX
XX  The invention relates to an isolated nucleic acid detection reagent
XX  capable of detecting 1000 or more genes from Drosophila. The invention is
XX  useful in developmental biology and in elucidating cell signalling and
XX  cell-cell interactions in higher eukaryotes for the development of
XX  insecticides, therapeutics and pharmaceutical drugs. The invention
XX  discloses genomic DNA sequences (ABL01840-ABL16175) and the encoded proteins.
XX  Sequences (ABL01840-ABL16175) and the encoded proteins.
XX  (ABB57737-ABB72072).
XX  The sequence data for this patent did not form part of the printed
XX  specification, but was obtained in electronic format directly from WIPO
XX  at ftp.wipo.int/pub/published_pct_sequences.
XX
XX  Sequence 968 AA;
SQ
Query Match 2.8%; Score 9; DB 22; Length 968;
Best Local Similarity 100.0%; Pred. No. 7.2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 217 ATVAGSTVT 225
DB 877 ATVAGSTVT 885
RESULT 4
AAG90076
ID AAG90076 standard; Protein; 223 AA.
XX

```

AC AAG90076;
XX 26-SEP-2001 (first entry)
XX C glutamicum protein fragment SEQ ID NO: 3830.
DE
XX Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
XX organic acid synthesis.
XX Corynebacterium glutamicum.
OS
XX EPI108790-A2.
XX
XX 20-JUN-2001.
PD
XX 18-DEC-2000; 2000EP-0127688.
XX
XX 16-DEC-1999; 99JP-0377484.
XX 07-APR-2000; 2000JP-0159162.
XX 03-AUG-2000; 2000JP-0280988.
XX
XX (KYOW) KYOWA HAKKO KOGYO KK.
XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX WPI; 2001-376931/40.
XX N-PSDB; AAH65295.
XX
XX Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
PI
XX
XX Claim 17; SEQ ID NO: 3830; 246pp + Sequence Listing; English.
XX
XX The present invention provides a number of nucleotide and protein
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of coryneform bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Coryneform bacterium, and identifying a homolog of a gene derived
CC from coryneform bacterium. Coryneform bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a protein described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
XX Sequence 223 AA;
SQ
Query Match 2.5%; Score 8; DB 22; Length 223;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 135 HTSPVAV 142
DB 130 HTSPVAV 137
RESULT 5
AAW94997
ID AAW94997 standard; Protein; 557 AA.
XX
XX AAW94997;
AC
XX 12-MAY-1999 (first entry)
DT
XX YAK-1 related serine/threonine protein kinase-HTLAR33.
DE
XX YAK-1; serine-threonine protein kinase; HTLAR33; bone loss; ARDS;
XX inflammatory disorder; osteoporosis; Adult Respiratory Disease Syndrome;
KM
DT

KW arthritis; psoriasis; dermatitis; asthma; allergy; infection; HIV-1;
KW HIV-2; cachexia; immunodeficient disorder; septic shock; pain; injury;
KW cancer; anorexia; bulimia; Parkinson's disease; cardiovascular disease;
KW neurological disorder; Huntington's disease; gene therapy; gene mapping;
KW Gilles de la Tourette's syndrome.
XX
XX Homo sapiens.
XX
XX EP94863-A1.
XX
XX 03-FEB-1999.
PD
XX 21-JUL-1998; 98EP-0305794.
XX
XX 20-FEB-1998; 98US-0027064.
XX 28-JUL-1997; 97US-0053924.
XX
XX (SMIK) SMITHKLINE BEECHAM CORP.
XX
XX Bergsma DJ, Shabon U;
PI WPI; 1999-108353/10.
XX N-PSDB; AAX17774.
XX
XX New serine-threonine kinase (HTLAR33) polypeptides and
PT polynucleotides - useful as diagnostic reagents and for prevention
PT and treatment of bone loss, neurological and inflammatory disorders
PT and cancer, HIV infections and angina pectoris
XX
XX
XX Claim 1; Page 15; 30pp; English.
XX
XX This represents a YAK-1 related serine-threonine protein kinase
CC polypeptide (HTLAR33). Host cells transformed with an expression vector
CC comprising the HTLAR33 nucleic acid are used for the recombinant
CC production of the protein. HTLAR33 polynucleotides and polypeptides are
CC useful for are useful for diagnosing susceptibility to diseases and for
CC screening for antagonists, agonists. These can be used in treatment to
CC enhance (agonist) or block (antagonist) HTLAR33 activity. Diseases
CC diagnosed, prevented or treated include bone loss and inflammatory
CC disorders, including osteoporosis, Adult Respiratory Disease Syndrome
CC (ARDS), Rheumatoid arthritis, osteoarthritis, inflammatory bowel disease
CC (IBD), psoriasis, dermatitis, asthma, allergies; bacterial, fungal,
CC protozoan and viral infections, especially those caused by HIV-1 or
CC HIV-2; HIV-associated cachexia and other immunodeficient disorders;
CC septic shock; pain; injury; cancers; anorexia; bulimia; Parkinson's
CC disease; cardiovascular disease including restenosis, atherosclerosis,
CC acute heart failure, myocardial infarction; hypotension; hypertension;
CC urinary retention; angina pectoris; ulcers; benign prostatic hypertrophy;
CC and psychotic and neurological disorders, including schizophrenia, manic
CC depression, anxiety; delirium, dementia, severe mental retardation and
CC dyskinesias, such as Huntington's disease or Gilles de la Tourette's
CC syndrome. HTLAR33 polypeptides are also useful for mapping genes to
CC chromosomes, allowing gene inheritance to be studied through linkage
CC analysis.
XX
XX Sequence 557 AA;
SQ
Query Match 2.5%; Score 8; DB 20; Length 557;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 179 APVAPPVT 186
DB 450 APVAPPVT 457
RESULT 6
AAW31599
ID AAW31599 standard; Peptide; 13 AA.
XX
XX AAW31599;
AC
XX 23-MAR-1998 (first entry)
DT


```

XX Linker peptide used in the construction of pIX38.6-K/L/H.
DE retroviral packaging system: human; viron protein; expression vector;
XX PCR primer; recombinant; linker peptide.
XX Synthetic.
OS Homo sapiens.
XX WO9707225-A2..
XX 27-FEB-1997.
XX 21-AUG-1996; 96WO-US13737.
XX 21-AUG-1995; 95US-0517488.
XX (CELL-) CELL GENESYS INC.
XX Cooke K, Dull TV, Farson DA, Finer MH, Zsebo KM;
XX WPI; 1997-165307/15.
XX High efficiency retroviral packaging system - used to transduce
XX human cells, esp. haematopoietic stem cells, T or B cells with
XX foreign genes
XX Disclosure; Page 27; 157pp; English.
XX This linker peptide is used to construct a plasmid pIX38.6-K/L/H which is
XX used in the construction of a retroviral packaging plasmid used for the
XX production of high titres of recombinant retrovirus in human cells. The
XX retroviral packaging plasmid comprises a retroviral helper DNA sequence
XX derived from a replication-incompetent retroviral genome that encodes,
XX in trans, all viron proteins required for packaging such a retroviral
XX vector. The helper DNA sequence encodes a ecotropic Moloney murine
XX leukaemia virus (MLV) or gibbon ape leukaemia virus (GALV) gag and pol;
XX and a xenotropic, amphotropic, ecotropic or polytropic envelope protein.
XX The packaging plasmids, designated KAT plasmids, are used with a second
XX retroviral vector encoding a foreign gene of interest to produce
XX mammalian cells with retroviral supernatants that express, e.g. a
XX hormone, lymphokine, growth factor or coagulation factor. The plasmids
XX are useful in construction of stable cell lines that constitutively
XX produce the retroviral proteins required in trans for the production of
XX retrovirus particles: gag, pol and env. The stable producer cells
XX continue to produce high titre retrovirus indefinitely in the absence of
XX drug selection due to the stable integration of both packaging function
XX and virus vector. The retroviral vector plasmids are constructed with
XX sequences enabling the episomal persistence without the need for stable
XX integration of the vector plasmid. The Env gene determines the host
XX range.
XX Sequence 13 AA;
SQ
Query Match 2.2%; Score 7; DB 18; Length 13;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 41 STSGSGS 47
DB 2 STSGSGS 8
RESULT 7
AAU11448
ID AAU11448 standard; Peptide; 13 AA.
XX AC AAU11448;
XX 12-MAR-2002 (first entry)
XX Synthetic peptide linker for chimaeric antibodies #1.
XX

```

```

KW Virucidal; cytostatic; peptide linker; anti-HIV;
KW human immunodeficiency virus infection; viral disease; malignant disease;
KW T cell; single chain antibody; cytomegalovirus; hepatitis C; hepatitis B;
KW mycobacterium avium.
XX Synthetic.
OS US6319494-B1.
XX 20-NOV-2001.
XX 07-JUN-1995; 95US-0479737.
XX 05-MAY-1994; 94US-0238405.
XX 14-DEC-1990; 90US-0627643.
XX 12-DEC-1991; 91WO-US09431.
XX 09-DEC-1992; 92US-0988194.
XX (CELL-) CELL GENESYS INC.
XX Capon DJ, Weiss A, Irving BA, Roberts MR, Zsebo K;
XX WPI; 2002-074399/10.
XX Treating viral infections, e.g. HIV, and malignancies using T cells
XX that express proteins which bind virus/tumour antigens and kill cells
XX presenting the antigens, via the activity of a cytoplasmic signal
XX transducing domain -
XX Example 3; Column 23; 38pp; English.
XX The invention relates to treating viral or malignant diseases using
XX modified T cells that express proteins comprising (in an N-terminal to
XX C-terminal direction) single chain antibody binding domains (that bind to
XX viral or tumour antigens), transmembrane domains and cytoplasmic signal
XX transducing domains. When the single chain antibody domain binds to the
XX viral or tumour antigen on the cell, the modified T cells kill cells
XX expressing the antigens. The modified T cells are used for treating
XX malignant diseases and viral infection especially by human
XX immunodeficiency virus (HIV), cytomegalovirus, hepatitis C, hepatitis B,
XX and mycobacterium avium. The present sequence is peptide linker used to
XX link together the functional peptides of the chimaeric antibody molecule
XX of the invention.
XX Sequence 13 AA;
SQ
Query Match 2.2%; Score 7; DB 23; Length 13;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 41 STSGSGS 47
DB 2 STSGSGS 8
RESULT 8
AAE10936
ID AAE10936 standard; peptide; 35 AA.
XX AC AAE10936;
XX 18-DEC-2001 (first entry)
XX Mouse masel protein conserved amino acid repeat motif #11.
XX Mouse; monogenic audiogenic seizure-susceptible-1 gene; masel;
XX transgenic animal; genetic abnormality; seizure.
XX Mus musculus.
XX WO200165927-A1.
XX 13-SEP-2001.
XX

```

XX	02-MAR-2001; 2001WO-US06962.
PF	
XX	03-MAR-2000; 2000US-187209P.
PR	03-AUG-2000; 2000US-222898P.
XX	
PA	(UTAH) UNIV UTAH RES FOUND.
XX	
PI	Ptacek L, White S, Fu Y, Skradaki S;
XX	WPI, 2001-589903/66.
DR	
PT	Monogenic Androgenic Seizure-Susceptible-1 (mass1) genes, useful for
FI	producing animal models of androgenic seizures -
PS	Disclosure; Fig 7, 79pp; English.
XX	
CC	The present invention relates to nucleic acid molecules designated
CC	monogenic androgenic seizure-susceptible-1 (mass1) genes. The nucleic
CC	acid molecule may be used via recombinant DNA methodologies in the
CC	production of transgenic animal (especially mouse) models for studying
CC	genetic abnormalities related to mass1 which result in seizure
CC	susceptible phenotypes (mass1 is androgenic seizures). The present
CC	sequence is mouse monogenic androgenic seizure-susceptible (mass1)
CC	protein conserved amino acid repeat motif.
CC	Note: The present sequence (SEQ ID NO:16) is incorrectly referred to as
CC	SEQ ID NO:15 in the sequence listing of the specification.
SQ	
Sequence	35 AA;
Query Match	2.2%; Score 7; DB 22; Length 35;
Best Local Similarity	100.0%; Pred. No. 31;
Matches	7; Conservative 0; Mismatches 0; Indels 0; Gaps 0
OY	287 TGORIAS 293 7 TGORIAS 13
Dbl	
RESULT 9	
ABP04087	ID ABP04087 standard; Protein, 54 AA.
XX	
AC	ABP04087;
XX	
DT	24-JUN-2002 (first entry)
DE	
XX	Human ORFX protein sequence SEQ ID NO:8156.
XX	
KW	Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
KW	hyperproliferative disorder; psoriasis; benign tumor; haemorrhage;
KW	dysregenerative disorder; osteoarthritis; neurodegenerative disorder;
KW	cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
KW	hypertension; hypothyroidism; cholesterol ester storage disease;
KW	immune deficiency; immune disorder; infectious disease;
KW	autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
KW	myasthenia gravis.
OS	Homo sapiens.
XX	
PN	WO200192523-A2.
XX	
PD	06-DEC-2001.
XX	
XX	29-MAY-2001; 2001WO-US10836.
PF	
XX	30-MAY-2000; 2000US-206132P.
PR	29-AUG-2000; 2000US-228716P.
XX	
PA	(CURA-) CURAGEN CORP.
XX	
PI	Shinkets RA, Leach MD;
XX	

DR WPI : 2002-106308/14.
XX
XX Novel human polypeptides and polynucleotides useful for diagnosing,
PT preventing and treating cardiovascular disease, neurodegenerative,
PT hyperproliferative disorders and autoimmune disorders -
XX
PS Disclosure; SEQ ID 8156; 1037bp; English.

The present invention describes substantially purified human proteins
CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1
CC in the specification). ABN15762 to ABN27252 encode the human ORFX
CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for
CC treating or preventing a pathology associated with an ORFX-associated
CC disorder in humans and in the manufacture of a medicament for treating a
CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide
CC sequences can be used in gene therapy. ORFX sequences can be used in the
CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
CC psoriasis, benign tumors, keloid, degenerative disorders, haemorrhage,
CC osteoarthritis, neurodegenerative diseases, diabetes mellitus, systemic
CC transplantation, cardiovascular diseases, diabetes mellitus, systemic
CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
CC storage disease, various immune deficiencyes and disorders, infectious
CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
CC disease and autoimmune inflammatory eye disease. ORFX proteins are also
CC useful for treating burns, incisions, ulcers, for treating osteoporosis,
CC bone degenerative disorders, or periodontal disease, and for gut
CC protection or regeneration and treatment of lung or liver fibrosis,
CC reperfusion injury in various tissues and conditions resulting from
CC systemic cytokine damage.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIDO
CC at ftp.wido.int/pub/published_pct_sequences.

SO Sequence 54 AA:

Query Match 2.2%; Score 7; DB 23; Length 54;
Beef Local Similarity 10.0%; Pred. No. 46;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0

QY 178 PAPVAPP 184
|||
Db 4 PAPVAPP 10

RESULT 10
AAM83197
ID AAM83197 standard; Protein; 66 AA.
XX
AC AAM83197;
XX
DT 07-NOV-2001 (first entry)
DE
XX Human immune/haematopoietic antigen SEQ ID NO:10750.
XX
KW Human; Immune; haematopoietic; immune/haematopoietic antigen; cancer;
CYC cytostatic; gene therapy; vaccine; metastasis.
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
PF 17-JAN-2001; 2001WO-US01354.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.

CC activity, and can be used in gene therapy and vaccine production. (1)
 CC proteins and polynucleotides may be used in the prevention, diagnosis and
 CC treatment of diseases associated with inappropriate (1) expression. For
 CC example, they may be used to treat disorders associated with decreased
 CC expression by rectifying mutations or deletions in a patient's genome
 CC that affect the activity of (1) by expressing inactive proteins or to
 CC supplement the patient's own production of (1). Additionally, (1)
 CC polynucleotides may be used to produce the secreted (1), by inserting the
 CC the nucleic acids into a host cell and culturing the cell to express the
 CC protein. (1) proteins and polynucleotides may be used to prevent,
 CC diagnose and treat immune/haematopoietic-related diseases, especially
 CC cancers and cancer metastases of haematopoietic-derived cells. AAK6703
 CC to AAK67694 represent human immune/haematopoietic antigen genomic
 CC sequences from the present invention. AAK54942 to AAK54950 and AAM62169
 CC represent sequences used in the exemplification of the present invention.

XX Sequence 66 AA;

Query Match 2.2%; Score 7; DB 22; Length 66;

Best Local Similarity 100.0%; Pred. No. 56; Mismatches 0; Indels 0; Gaps 0;

Matches 7; Conservative 0;

166 KKPTPTP 172

11 KKPTPTP 17

Db

RESULT 11
 AAY41355 standard; Protein; 67 AA.

XX AAY41355;

DT 02-DEC-1999 (first entry)

XX Human secreted protein encoded by gene 48 clone HMAAP70.

XX Human; secreted protein; fusion protein; gene therapy; protein therapy;
 XX diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukemia;
 XX developmental abnormality; foetal deficiency; blood; allergy; renal;
 XX immune system; aschima; lymphocytic disease; brain; hepatic; lymphoma;
 XX inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
 XX cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
 XX osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
 XX endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.

XX Homo sapiens.

XX MO9947540-A1.

XX 23-SEP-1999.

XX 18-MAR-1999; 99WO-US05804.

XX 19-MAR-1998; 98US-0078563.

XX 19-MAR-1998; 98US-0078566.

XX 19-MAR-1998; 98US-0078573.

XX 19-MAR-1998; 98US-0078574.

XX 19-MAR-1998; 98US-0078576.

XX 19-MAR-1998; 98US-0078577.

XX 19-MAR-1998; 98US-0078578.

XX 19-MAR-1998; 98US-0078579.

XX 01-APR-1998; 98US-0080312.

XX 01-APR-1998; 98US-0080313.

XX 01-APR-1998; 98US-0080314.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Ruben SM, Ni J, Rosen CA, Yu G, Young PE, Feng P, Soppet DR;

XX Wei Y, Endress GA, Duan RD, Kyaw H, Ebner R, Laffleur DW;

XX Olsen HS, Shi Y, Moore PA;

DR WPI; 1999-562050/47.

XX N-PSDB; AA224858.

PT New isolated human genes, useful for diagnosis and treatment of e.g.

PT cancers, neurological disorders, immune diseases, inflammation or blood

PT disorders -

XX Claim 11; Page 387; 484p; English.

XX This sequence represents a secreted human protein encoded by the gene

XX clone detailed in the descriptor line. The gene can be used to generate

XX fusion proteins by linking to the gene to a human immunoglobulin Fc

XX portion (e.g. AA224802) for increasing the stability of the fused

XX protein (e.g. AA224802) for increasing the stability of the fused

XX The invention relates to the human protein only.

XX acid sequences: AA224811-224907; amino acid sequences AAY41308-Y41404)

XX which are useful for preventing, treating or ameliorating medical

XX conditions e.g. by protein or gene therapy. Also, pathological

XX polypeptides in a sample or by determining the amount of the new

XX the new polynucleotides. Specific uses are described for each of the 95

XX polynucleotides, based on which tissues they are most highly expressed in

XX (see AA224811 for described uses).

XX Sequence 67 AA;

Query Match 2.2%; Score 7; DB 20; Length 67;

Best Local Similarity 100.0%; Pred. No. 57; Mismatches 0; Indels 0; Gaps 0;

Matches 7; Conservative 0;

18 LGLIRGV 24

11 LGLIRGV 17

Db

RESULT 12
 AAU61587 standard; Protein; 72 AA.

XX AAU61587;

XX 27-FEB-2002 (first entry)

XX Propionibacterium acnes immunogenic protein #22483.

XX SAPHO syndrome; synovitis; acne; pustulosis; hyperostosis; osteomyelitis;

XX uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;

XX dermatological; osteopathic; neuroprotectant.

XX Propionibacterium acnes.

XX MO200181581-A2.

XX 01-NOV-2001.

XX 20-APR-2001; 2001WO-US12865.

XX 21-APR-2000; 2000US-199047P.

XX 02-JUN-2000; 2000US-209841P.

XX 07-JUL-2000; 2000US-216747P.

XX (CORI-) CORIXA CORP.

XX Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;

XX L'maisonmeuve J, Zhang Y, Jen S, Carter D;

XX WPI; 2001-616774/71.

XX N-PSDB; AAS59618.

PT Propionibacterium acnes polypeptides and nucleic acids useful for

PT vaccinating against and diagnosing infections, especially useful for

PT treating acne vulgaris -

XX Example 1, SEQ ID No 22782; 1069pp; English.
PS
XX
CC Sequences AAU39105-AAU68017 represent *Propionibacterium acnes* immunogenic
CC polypeptides. The proteins and their associated DNA sequences are used in
CC the treatment, prevention and diagnosis of medical conditions caused by
CC *P. acnes*. The disorders include SAPHO syndrome (synovitis, acne,
CC pustulosis, hyperostosis and osteomyelitis), uveitis and endophthalmitis.
CC *P. acnes* is also involved in infections of bone, joints and the central
CC nervous system, however it is particularly involved in the inflammatory
CC lesions associated with acne vulgaris. A method for detecting the
CC presence or absence of *P. acnes* in a patient comprises contacting a
CC sample with a binding agent that binds to the proteins of the invention
CC and determining the amount of bound protein in the sample. The
CC polypeptides may be used as antigens in the production of antibodies
CC specific for *P. acnes* proteins. These antibodies can be used to
CC downregulate expression and activity of *P. acnes* polypeptides and
CC therefore treat *P. acnes* infections. The antibodies may also be used as
CC diagnostic agents for determining *P. acnes* presence, for example, by
CC enzyme linked immunosorbent assay (ELISA).
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
SQ
XX
SQ Sequence 72 AA;
Query Match 2.2%; Score 7; DB 22; Length 72;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 178 PAPVAPP 184
|||
Db 28 PAPVAPP 34
RESULT 13
AAU50707
ID AAU50707 standard; Protein; 80 AA.
XX
AC AAU50707;
XX
DT 27-FEB-2002 (first entry)
XX
DE *Propionibacterium acnes* immunogenic protein #11603.
XX
KM SAPHO syndrome; synovitis; acne; pustulosis; hyperostosis; osteomyelitis;
KM uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KM inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KM dermatological; osteopathic; neuroprotectant.
XX
OS *Propionibacterium acnes*.
XX
PN WO200181581-A2.
XX
PD 01-NOV-2001.
XX
PE 20-APR-2001; 2001WO-US12865.
XX
PR 21-APR-2000; 2000US-199047P.
PR 02-JUN-2000; 2000US-208841P.
PR 07-JUL-2000; 2000US-216747P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Skeiky YAM, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX
DR WPI; 2001-616774/71.
DR N-PSDB; AAS59549.
XX
PT *Propionibacterium acnes* polypeptides and nucleic acids useful for
PT vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris -

XX Example 1, SEQ ID No 11902; 1069pp; English.
PS
XX
CC Sequences AAU39105-AAU68017 represent *Propionibacterium acnes* immunogenic
CC polypeptides. The proteins and their associated DNA sequences are used in
CC the treatment, prevention and diagnosis of medical conditions caused by
CC *P. acnes*. The disorders include SAPHO syndrome (synovitis, acne,
CC pustulosis, hyperostosis and osteomyelitis), uveitis and endophthalmitis.
CC *P. acnes* is also involved in infections of bone, joints and the central
CC nervous system, however it is particularly involved in the inflammatory
CC lesions associated with acne vulgaris. A method for detecting the
CC presence or absence of *P. acnes* in a patient comprises contacting a
CC sample with a binding agent that binds to the proteins of the invention
CC and determining the amount of bound protein in the sample. The
CC polypeptides may be used as antigens in the production of antibodies
CC specific for *P. acnes* proteins. These antibodies can be used to
CC downregulate expression and activity of *P. acnes* polypeptides and
CC therefore treat *P. acnes* infections. The antibodies may also be used as
CC diagnostic agents for determining *P. acnes* presence, for example, by
CC enzyme linked immunosorbent assay (ELISA).
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
SQ
XX
SQ Sequence 80 AA;
Query Match 2.2%; Score 7; DB 22; Length 80;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 43 SGSGSHR 49
|||
Db 43 SGSGSHR 49
RESULT 14
AA96225
ID AA96225 standard; Protein; 90 AA.
XX
AC AA96225;
XX
DT 11-SEP-2000 (first entry)
XX
DE *Arabidopsis* LEC1 consensus region.
XX
KM Leafy cotyledon 1 transcriptional activator; LEC1; apomixis;
KM selectable marker; transgenic plant; transgenic seed; HAP3.
XX
OS *Arabidopsis* sp.
XX
PN WO200028058-A2.
XX
PD 18-MAY-2000.
XX
PE 09-NOV-1999; 99WO-US26514.
XX
PR 09-NOV-1998; 98US-0107643.
PR 10-NOV-1998; 98US-0107810.
XX
PA (PION-) PIONEER HI-BRED INT INC.
PA (DUPO) DU PONT DE NEMOURS & CO E I.
XX
PI Lowe KS, Gordon-Kamm WJ, Klein TM, Rasco-Gaunt S, Cahoon RE;
PI Sun X, Hoerster GU, Gregory CA, Nadimpalli R;
XX
DR WPI; 2000-376568/32.
XX
PT New HAP3-type CCAAT-box binding transcriptional activators,
PT particularly leafy cotyledon 1 transcriptional activator, useful for
PT inducing somatic embryogenesis or apomixis in a plant cell -
XX
PS Example 6; Fig 1, 94pp; English.

The present sequence is the leafy cotyledon 1 transcriptional activator, LEC1 consensus sequence from Arabidopsis. This sequence is a HAP3-type CCAT-box binding protein. LEC1 expression initiates the formation of embryo-like structures and improves growth and recovery of transformants. When apomixis occurs, i.e. the replacement of sexual reproduction by asexual reproduction, LEC1 expression in the nucleus would trigger embryo formation from maternal tissues only. This results in the production of seeds identical to the parent. Using LEC1, transgenic high yielding seeds could be developed. In addition, LEC1 could be used for positive selection of a transformed cell (transgenic plant) for increasing transformation efficiency and for increasing recovery of regenerated plants. The present sequence was used in the identification of the plant LEC1 consensus sequence (AY96223).

Sequence 90 AA;

Query Match 2.2%; Score 7; DB 21; Length 90;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 314 YVDPPLTV 320
|||
78 YVDPPLTV 84

Db

RESULT 15
ABB55442
ID ABB55442 standard; Protein; 101 AA.

AC ABB55442;

DT 16-MAY-2002 (first entry)

XX Lactococcus lactis protein rplX.

XX Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.

OS Lactococcus lactis IL1403.

PN FR2807446-A1.

XX 12-OCT-2001.

XX 11-APR-2000; 2000FR-0004630.

XX 11-APR-2000; 2000FR-0004630.

PA (INRG) INRA INST NAT RECH AGRONOMIQUE.

PI Bolotline A, Sorokine A, Renault P, Ehrlich SD;

DR WPI; 2002-043418/06.

XX New nucleotide sequence useful in the identification of Lactococcus

PT lactic and related species -

PS Claim 6; SEQ ID No 2144; 2504bp; French.

The present invention is related to a Lactococcus lactis nucleotide sequence (ABA90521) and related proteins (ABB53300-ABB55621). The nucleic acid sequence is useful in the detection and/or amplification of related species. The proteins of the invention are useful for the biosynthesis or biodegradation of a composition of interest. The invention helps research in lactic bacteria, particularly useful in the production of yogurt and cheese.
Note: The sequence data for this patent is based on equivalent patent WO2001/7334 (published 18-OCT-2001) which is available in electronic format directly from WIPO at ftp.wipo.int/pub/published_pat_sequences.

Sequence 101 AA;

Query Match 2.2%; Score 7; DB 23; Length 101;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 279 VKTGDTV 285
|||
3 VKTGDTV 9

Db

RESULT 16

AG58104
ID AG58104 standard; Protein; 108 AA.

AC AG58104;

DT 18-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 74962.

XX Protein identification; signal transduction pathway; metabolic pathway;

XX hybridisation assay; genetic mapping; gene expression control; promoter;

XX termination sequence.

OS Arabidopsis thaliana.

PN EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

XX 05-MAR-1999; 99US-0123180.

XX 09-MAR-1999; 99US-0123548.

XX 23-MAR-1999; 99US-0125788.

XX 25-MAR-1999; 99US-0126264.

XX 29-MAR-1999; 99US-0126785.

XX 01-APR-1999; 99US-0127462.

XX 06-APR-1999; 99US-0128234.

XX 16-APR-1999; 99US-0128714.

XX 19-APR-1999; 99US-0129845.

XX 21-APR-1999; 99US-0130077.

XX 23-APR-1999; 99US-0130449.

XX 28-APR-1999; 99US-0130891.

XX 30-APR-1999; 99US-0131449.

XX 04-MAY-1999; 99US-0132048.

XX 05-MAY-1999; 99US-0132484.

XX 06-MAY-1999; 99US-0132485.

XX 07-MAY-1999; 99US-0132486.

XX 11-MAY-1999; 99US-0132487.

XX 14-MAY-1999; 99US-0132488.

XX 14-MAY-1999; 99US-0134219.

XX 18-MAY-1999; 99US-0134370.

XX 19-MAY-1999; 99US-0134768.

XX 20-MAY-1999; 99US-0135124.

XX 21-MAY-1999; 99US-0135353.

XX 24-MAY-1999; 99US-0135623.

XX 27-MAY-1999; 99US-0136021.

XX 28-MAY-1999; 99US-0136392.

XX 01-JUN-1999; 99US-0137222.

XX 03-JUN-1999; 99US-0137528.

XX 04-JUN-1999; 99US-0137502.

XX 07-JUN-1999; 99US-0137724.

XX 08-JUN-1999; 99US-0138094.

XX 10-JUN-1999; 99US-0138540.

XX 10-JUN-1999; 99US-0138847.

PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139753.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 19-JUL-1999; 99US-0144336.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145087.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145219.
PR 26-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.

PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151067.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154039.
PR 16-SEP-1999; 99US-0154018.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158079.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match

Best Local Similarity 2.2%; Score 7; DB 21; Length 108;

Matches 7; Conservativity 100.0%; Pred. No. 88; 0; Mismatches 0; Indels 0; Gaps 0;

QY 123 RERSISS 129

DB 99 RERSISS 105

```
RESULT 17
ABR66134
ID ABR66134 standard; Protein; 112 AA.
XX
XX ABR66134;
AC
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster polypeptide SEQ ID NO 25194.
XX
XX Drosophila developmental biology; cell signalling; insecticide;
XX pharmaceutical.
XX
XX Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001;
XX
XX 23-MAR-2001; 2001WO-US09231.
XX
XX 23-MAR-2000; 2000US-191637P.
XX
XX 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PMD, Myers EW;
XX
XX WPI; 2001-656860/75.
XX
XX N-PSDB; ABL10237.
XX
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX
XX Disclosure; SEQ ID NO 25194; 21pp + Sequence Listing; English.
XX
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABR57737-ABR12072).
XX
XX (ABR57737-ABR12072).
XX
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_sequences.
XX
XX
XX Sequence 112 AA;
SQ
XX
XX
XX Query Match 2.2%; Score 7; DB 22; Length 112;
XX Best Local Similarity 100.0%; Pred. No. 92;
XX Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 51 SSGGGLA 57
DB 96 SSGGGLA 102
XX
XX
XX RESULT 18
XX ABR00920
XX ID ABR00920 standard; Protein; 114 AA.
XX
XX
XX ABR00920;
XX
XX 24-JUN-2002 (first entry)
XX
XX Human ORFX protein sequence SEQ ID NO:1822.
XX
XX Human, open reading frame; ORFX; gene therapy; cancer; cirrhosis;
XX
```

```
XX hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
XX degenerative disorder; osteoarthritis; neurodegenerative disorder;
XX cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
XX hypertension; hypothyroidism; cholesterol ester storage disease;
XX immune deficiency; immune disorder; infectious disease;
XX autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
XX myasthenia gravis.
XX
XX Homo sapiens.
XX
XX WO200192523-A2.
XX
XX 06-DEC-2001.
XX
XX 29-MAY-2001; 2001WO-US10836.
XX
XX 30-MAY-2000; 2000US-206132P.
XX
XX 29-AUG-2000; 2000US-228716P.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Shinketsu RA, Leach MD;
XX
XX WPI; 2002-106308/14.
XX
XX N-PSDB; ABR16672.
XX
XX
XX Novel human polypeptides and polynucleotides useful for diagnosing,
XX preventing and treating cardiovascular disease, neurodegenerative,
XX hyperproliferative disorders and autoimmune disorders -
XX
XX Disclosure; SEQ ID 1822; 1037pp; English.
XX
XX
XX The present invention describes substantially purified human proteins
XX (referred to as open reading frame, ORFX, where x is 1-11491 (see Table 1
XX in the specification). ABR15762 to ABR27252 encode the human ORFX
XX proteins given in ABR00010 to ABR11500. ORFX proteins are useful for
XX treating or preventing a pathology associated with an ORFX-associated
XX disorder in humans, and in the manufacture of a medicament for treating a
XX syndrome associated with ORFX-associated disorder. ORFX polynucleotide
XX sequences can be used in gene therapy. ORFX sequences can be used in the
XX treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
XX psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
XX osteoarthritis, neurodegenerative disorders, disorders related to organ
XX transplantation, cardiovascular diseases, diabetes mellitus, systemic
XX lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
XX storage disease, various immune deficiencies and disorders, infectious
XX diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
XX arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
XX disease and autoimmune inflammatory eye disease. ORFX proteins are also
XX useful for treating burns, incisions, ulcers, for treating osteoporosis,
XX bone degenerative disorders, or periodontal disease, and for gut
XX protection or regeneration and treatment of lung or liver fibrosis,
XX systemic cytokine damage.
XX
XX N.B. The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_sequences.
XX
XX
XX Sequence 114 AA;
SQ
XX
XX
XX Query Match 2.2%; Score 7; DB 23; Length 114;
XX Best Local Similarity 100.0%; Pred. No. 93;
XX Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 99 NNINNSY 105
DB 64 NNINNSY 70
XX
XX
XX RESULT 19
XX ABR04226
XX ID ABR04226 standard; Protein; 126 AA.
XX
XX
```


AC ABG04226;
 XX
 XX 13-FEB-2002 (first entry)
 XX
 DE Novel human diagnostic protein #4217.
 XX
 KM Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KM food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 DR WPI; 2001-639362/73.
 DR N-PSDB; AAS68413.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX
 PS Claim 20; SEQ ID NO 34585; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX
 SQ Sequence 126 AA;
 QY
 Db 181 VAPPVTE 187
 112 VAPPVTE 118
 Query Match 2.2%; Score 7; DB 22; Length 126;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 20
 ID AAO05078
 AC AAO05078; standard; Protein; 136 AA.
 AC AAO05078;
 XX

DT 06-NOV-2001 (first entry)
 XX
 DE Human polypeptide SEQ ID NO 18970.
 XX
 KM Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KM vaccine; peptide therapy; stem cell growth factor; hematopoiesis;
 KM tissue growth factor; immunomodulatory; cancer; leukemia;
 KM nervous system disorders; arthritis; inflammation.
 XX
 OS Homo sapiens.
 XX
 PN WO200164835-A2.
 XX
 PD 07-SEP-2001.
 XX
 PF 26-FEB-2001; 2001WO-US04927.
 XX
 PR 28-FEB-2000; 2000US-0515126.
 PR 18-MAY-2000; 2000US-0577409.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Drmanac RT;
 XX
 DR WPI; 2001-514838/56.
 DR N-PSDB; AAI85009.
 XX
 PT Isolated nucleic acids and polypeptides, useful for preventing
 PT diagnosing and treating e.g. leukemia, inflammation and immune
 PT disorders -
 XX
 PS Claim 20; SEQ ID NO 18970; 1399pp + Sequence Listing; English.
 XX
 CC The invention relates to human polynucleotides (AAI79941-AAI93841) and
 CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, hematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activity/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukemia, nervous system disorders, arthritis and
 CC inflammation.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX
 SQ Sequence 136 AA;
 QY
 Db 167 KPTPTTP 173
 28 KPTPTTP 34
 Query Match 2.2%; Score 7; DB 22; Length 136;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 RESULT 21
 ID AAG58103
 AC AAG58103; standard; Protein; 138 AA.
 AC AAG58103;
 XX
 DT 18-OCT-2000 (first entry)
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 74961.
 XX
 KM Protein identification; signal transduction pathway; metabolic pathway;
 KM hybridisation assay; genetic mapping; gene expression control; promoter;
 KM termination sequence.
 XX


```

PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160960.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

Query Match 2.2%; Score 7; DB 21; Length 138;
 Best Local Similarity 100.0%; Pred.No.1.1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 123 RERSISS 129
 |||||
 Db 129 RERSISS 135

RESULT 22

```

AA005875 ID AA005875 standard; Protein; 147 AA.
XX AC AA005875;
XX DT 06-NOV-2001 (first entry)
XX DE Human polypeptide SEQ ID NO 19767.
XX KM Human; cytokine; cell proliferation; cell differentiation; gene therapy;
XX KM tissue vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
XX KM tissue growth factor; immunomodulatory; cancer; leukaemia;
XX KM nervous system disorders; arthritis; inflammation.
XX OS Homo sapiens.
XX PN WO200164835-A2.
XX PD 07-SEP-2001.
XX
```

```

PF 26-FEB-2001; 2001WO-US04927.
XX 28-FEB-2000; 2000US-0515126.
PR 18-MAY-2000; 2000US-0577409.
XX (HYSE-) HYSEQ INC.
XX PI Tang YT, Liu C, Drmanac RT;
XX WPI, 2001-514836/56.
XX N-PSDB; AA185806.
XX Isolated nucleic acids and polypeptides, useful for preventing
XX diagnosing and treating e.g. leukaemia, inflammation and immune
XX disorders.
XX Claim 20; SEQ ID NO 19767; 1399pp + Sequence Listing; English.
XX The invention relates to human polynucleotides (AA179941-AA193841) and
XX the encoded proteins (AA000010-AA013910) that exhibit activity elating to
XX cytokine, cell proliferation or cell differentiation or which may induce
XX production of other cytokines in other cell populations. The
XX polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX peptide therapy. The polypeptides have various cytokine-like activities,
XX e.g. stem cell growth factor activity, haematopoiesis regulating
XX activity, tissue growth factor activity, immunomodulatory activity and
XX activity/inhibit activity and may be useful in the diagnosis and/or
XX treatment of cancer, leukaemia, nervous system disorders, arthritis and
XX inflammation.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 147 AA;

```

Query Match 2.2%; Score 7; DB 22; Length 147;
 Best Local Similarity 100.0%; Pred.No.1.2e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 178 PAPVAPP 184
 |||||
 Db 25 PAPVAPP 31

RESULT 23

```

AAE02227 ID AAE02227 standard; Protein; 157 AA.
XX AC AAE02227;
XX DT 31-JUL-2001 (first entry)
XX DE Staphylococcus aureus biofilm protein, biofilm A.
XX KM Biofilm A; antibacterial; bacteraemia; food poisoning; osteomyelitis;
XX KM otitis media; tracheitis; epiglottitis; chryoiditis; empyema; abscess;
XX KM lung; splenic; retroperitoneal; endocarditis; cerebral; renal; diarrhoea;
XX KM blepharitis; conjunctivitis; keratitis; endophthalmitis; dacryocystitis;
XX KM cellulitis; epididymitis; impetigo; folliculitis; wound; myositis;
XX KM septic arthritis; therapy.
XX OS Staphylococcus aureus 'WCUH 29'.
XX PN WO200130806-A1.
XX PD 03-MAY-2001.
XX PF 26-OCT-2000; 2000WO-US29460.
XX PR 28-OCT-1999; 99US-0161984.
XX (SMIK ) SMITHKLINE BEECHAM CORP.
XX (SMIK ) SMITHKLINE BEECHAM PLC.
XX PA
```

XX Bryant J, Burnham MKR, Lunsford RD, Throup JP;
PI WPI: 2001-308606/32.
XX N-PSDB; AAD05669.
XX New biofilm protein from *Staphylococcus*, useful e.g. in screening for
PT antibacterial agents, and related nucleic acid
PS Claim 1; Page 5; 66pp; English.

CC The present sequence is *Staphylococcus aureus* WCUH 29 strain biofilm
CC protein, biofilm A. The biofilm polypeptides are useful for the
CC diagnosis and treatment of bacterial infections specifically caused by
CC *Staphylococcus*, especially associated with formation of biofilm in a
CC wound or on an (in-dwelling) medical device. The polypeptide is used in
CC the treatment of diseases such as invasive and toxicogenic diseases,
CC bacteraemia in cancer patients. It is also used in the treatment of
CC diseases caused by the action of exotoxins, including food poisoning,
CC scalded skin syndrome, infections of upper respiratory tract (otitis
CC media, bacterial tracheitis, acute epiglottitis, thyroiditis), lower
CC respiratory (empyema, lung abscess), cardiac abscess, retroperitoneal
CC gastrointestinal (secretory diarrhoea, splenic abscess, conjunctivitis,
CC keratitis, endophthalmitis, preseptal and orbital cellulitis,
CC dacryocystitis), kidney and urinary tract (epididymitis, intrarenal
CC and perinephric abscess, toxic shock syndrome), skin (impetigo,
CC folliculitis, cutaneous abscess, cellulitis, wound infection, bacterial
CC myositis) bone and joint (septic arthritis, osteomyelitis).

XX Sequence 157 AA;
SQ

Query Match 2.2%; Score 7; DB 22; Length 157;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 LGLIFGV 24
DB 79 LGLIFGV 85

RESULT 24
AAU66705
ID AAU66705 standard; Protein; 159 AA.
XX
AC AAU66705;
XX
DT 13-FEB-2002 (first entry)
XX
DE Propionibacterium acnes immunogenic protein #27601.
XX
KW SAPHO syndrome; synovitis; acne; pustulosis; hyperostosis; osteomyelitis;
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW dermatological; osteopathic; neuroprotectant.
XX
OS Propionibacterium acnes.
XX
XX WO200181581-A2.
XX
XX 01-NOV-2001.
XX
XX 20-APR-2001; 2001WO-US12865.
XX
XX 21-APR-2000; 2000US-199047P.
XX
XX 02-JUN-2000; 2000US-208841P.
XX
XX 07-JUL-2000; 2000US-216747P.
XX
XX (CORI-) CORIXA CORP.
XX
XX Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
XX L'maisonmeuve J, Zhang Y, Jen S, Carter D;

DR WPI: 2001-616774/71.
DR N-PSDB; AAS59749.
XX
XX Propionibacterium acnes polypeptides and nucleic acids useful for
PT vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris
XX
XX Example 1; SEQ ID NO 27900; 1069pp; English.

CC Sequences AAU39105-AAU68017 represent *Propionibacterium acnes* immunogenic
CC polypeptides. The proteins and their associated DNA sequences are used in
CC the treatment, prevention and diagnosis of medical conditions caused by
CC *P. acnes*. The disorders include SAPHO syndrome (synovitis, acne,
CC pustulosis, hyperostosis and osteomyelitis), uveitis and endophthalmitis.
CC *P. acnes* is also involved in infections of bone, joints and the central
CC nervous system, however it is particularly involved in the inflammatory
CC lesions associated with acne vulgaris. A method for detecting the
CC presence or absence of *P. acnes* in a patient comprises contacting a
CC sample with a binding agent that binds to the proteins of the invention
CC and determining the amount of bound protein in the sample. The
CC polypeptides may be used as antigens in the production of antibodies
CC specific for *P. acnes* proteins. These antibodies can be used to
CC downregulate expression and activity of *P. acnes* polypeptides and
CC therefore treat *P. acnes* infections. The antibodies may also be used as
CC diagnostic agents for determining *P. acnes* presence, for example, by
CC enzyme linked immunosorbent assay (ELISA).
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 159 AA;
SQ

Query Match 2.2%; Score 7; DB 22; Length 159;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 162 VVVVKKP 168
DB 151 VVVVKKP 157

RESULT 25
AAU43362
ID AAU43362 standard; Protein; 159 AA.
XX
AC AAU43362;
XX
DT 27-FEB-2002 (first entry)
XX
DE Propionibacterium acnes immunogenic protein #4258.
XX
KW SAPHO syndrome; synovitis; acne; pustulosis; hyperostosis; osteomyelitis;
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW dermatological; osteopathic; neuroprotectant.
XX
OS Propionibacterium acnes.
XX
XX WO200181581-A2.
XX
XX 01-NOV-2001.
XX
XX 20-APR-2001; 2001WO-US12865.
XX
XX 21-APR-2000; 2000US-199047P.
XX
XX 02-JUN-2000; 2000US-208841P.
XX
XX 07-JUL-2000; 2000US-216747P.
XX
XX (CORI-) CORIXA CORP.
XX
XX Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
XX L'maisonmeuve J, Zhang Y, Jen S, Carter D;

DR WPI; 2001-616774/71.
 DR N-PSDB; AAS59520.
 XX
 PT Propionibacterium acnes polypeptides and nucleic acids useful for
 PT vaccinating against and diagnosing infections, especially useful for
 PT treating acne vulgaris -
 XX
 PS Example 1; SEQ ID No 4557; 1069pp; English.
 XX
 CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
 CC polypeptides. The proteins and their associated DNA sequences are used in
 CC the treatment, prevention and diagnosis of medical conditions caused by
 CC P. acnes. The disorders include SAPRO syndrome (synovitis, acne,
 CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
 CC P. acnes is also involved in infections of bone, joints and the central
 CC nervous system, however it is particularly involved in the inflammatory
 CC lesions associated with acne vulgaris. A method for detecting the
 CC presence or absence of P. acnes in a patient comprises contacting a
 CC sample with a binding agent that binds to the proteins of the invention
 CC and determining the amount of bound protein in the sample. The
 CC polypeptides may be used as antigens in the production of antibodies
 CC specific for P. acnes proteins. These antibodies can be used to
 CC downregulate expression and activity of P. acnes polypeptides and
 CC therefore treat P. acnes infections. The antibodies may also be used as
 CC diagnostic agents for determining P. acnes presence, for example, by
 CC enzyme linked immunosorbent assay (ELISA).
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 SQ Sequence 159 AA;
 XX
 QY Query Match 2.2%; Score 7; DB 22; Length 159;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 XX
 Db 162 VVVVKKP 168
 151 VVVVKKP 157
 XX
 RESULT 26
 ABP25478
 ID ABP25478 standard; Protein; 179 AA.
 AC
 XX ABP25478;
 DT 02-JUL-2002 (first entry)
 XX
 DE Streptococcus polypeptide SEQ ID NO 132.
 XX
 KM Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
 KM group A streptococcus; Streptococcus pyogenes; antibacterial;
 KM antiInflammatory; infection; vaccine; meningitis; gene therapy.
 XX
 OS Streptococcus pyogenes.
 XX
 PN WO200234771-A2.
 PD 02-MAY-2002.
 XX
 PF 29-OCT-2001; 2001MO-GB04789.
 XX
 PR 27-OCT-2000; 2000GB-0026333.
 PR 24-NOV-2000; 2000GB-0028727.
 PR 07-MAR-2001; 2001GB-0005640.
 XX
 PA (CHIR-) CHIRON SPA.
 PA (GENO-) INST GENOMIC RES.
 XX
 PI Telford J, Maignani V, Margarit Ros YI, Grandi G, Frazer C;
 PI Tetzelin H;
 XX

DR WPI; 2002-352536/38.
 DR N-PSDB; AEN66109.
 XX
 PT New Streptococcus protein for the treatment or prevention of infection
 PT or disease caused by Streptococcus bacteria, such as meningitis, and
 PT for detecting a compound that binds to the protein -
 XX
 PS Claim 1; Page 3167; 4525pp; English.
 XX
 CC The invention relates to a protein (ABP25413-ABP30895) from group B
 CC streptococcus/GAS (Streptococcus agalactiae) or group A streptococcus/GAS
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
 CC the specification. The proteins have antibacterial and antiinflammatory
 CC activity. (I), nucleic acids encoding (I), AEN66044-ABN71526 and
 CC antibodies that bind (I) are used in the manufacture of medicaments for
 CC the treatment or prevention of infection or disease caused by
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
 CC Nucleic acids encoding (I) are used to detect Streptococcus in a
 CC biological sample. (I) is used to determine whether a compound binds to
 CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
 CC used as a vaccine or diagnostic composition. The disease caused by
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic
 CC acid encoding (I) may be used to recombinantly produce (I) and may be
 CC used in gene therapy. Antibodies to (I) are used for affinity
 CC chromatography, immunoassays, and distinguishing/identifying
 CC Streptococcus proteins.
 CC
 SQ Sequence 179 AA;
 XX
 QY Query Match 2.2%; Score 7; DB 23; Length 179;
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 XX
 Db 52 GSGGIAL 58
 51 GSGGIAL 57
 XX
 RESULT 27
 AAY75693
 ID AAY75693 standard; Protein; 185 AA.
 AC
 XX AAY75693;
 DT 21-MAR-2000 (first entry)
 XX
 DE Neisseria gonorrhoeae ORF 931 protein sequence SEQ ID NO:2858.
 XX
 KM Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
 KM antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
 KM antibacterial; gene therapy.
 XX
 OS Neisseria gonorrhoeae.
 XX
 PN WO9957280-A2.
 PD 11-NOV-1999.
 XX
 PF 30-APR-1999; 99WO-US09346.
 XX
 PR 01-MAY-1998; 98US-0083758.
 PR 31-JUL-1998; 98US-0094869.
 PR 02-SEP-1998; 98US-0098994.
 PR 02-SEP-1998; 98US-0099062.
 PR 09-OCT-1998; 98US-0103749.
 PR 09-OCT-1998; 98US-0103794.
 PR 09-OCT-1998; 98US-0103796.
 PR 25-FEB-1999; 99US-0121528.
 XX
 PA (CHIR) CHIRON CORP.
 PA (GENO-) INST GENOMIC RES.
 XX
 PI Fraser C, Galeotti C, Grandi G, Hickey E, Maignani V, Mora M;
 PI

PI Petersen J, Pizsa M, Rappuoli R, Ratti G, Scalato E, Scarsella M;
PI Tettelin H, Venter JC;
XX
XX WPI: 2000-062150/05.
DR N-PSDB; AAZ54455.
XX

PT Novel Neisseria polypeptides predicted to be useful antigens for
PT vaccines and diagnostics -
XX
XX
XX

PS Claim 2; Page 1343; 1453pp; English.

XX AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAV74253 to AAV75941
CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ5473 represent
CC PCR primers used in the exemplification of the present invention. The
CC polypeptides, the polynucleotides, antibodies and compositions of
CC the invention can be used as vaccines, as diagnostic reagents, and as
CC immunogenic compositions. The polypeptides can be used in the
CC manufacture of medicaments for treating or preventing infection due to
CC Neisseria bacteria (e.g. meningitis and septicemia), to detect the
CC presence of Neisseria bacteria, or to raise antibodies. They may also
CC be used to screen for agonists or antagonists, which may themselves
CC have use as antibacterial agents. The polynucleotides of the invention
CC may also be used in gene therapy protocols.

XX Sequence 185 AA;

Query Match 2.2%; Score 7; DB 21; Length 185;

Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 79 DTVSKIA 85
|||
152 DTVSKIA 158

RESULT 28

ABBE1524 ID ABBE1524 standard; Protein; 195 AA.

XX ABBE1524;

XX 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 11364.

XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.

XX Drosophila melanogaster.

OS WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI: 2001-656660/75.

XX N-PSDB; ABL05627.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX

PS Disclosure; SEQ ID NO 11364; 21pp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
CC sequences (AB101840-AB16175) and the encoded proteins
CC (AAB57737-AB572072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 195 AA;

Query Match 2.2%; Score 7; DB 22; Length 195;

Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 51 SGGGCLA 57
|||
179 SGGGCLA 185

RESULT 29

AAW71722 ID AAW71722 standard; Protein; 208 AA.

XX AAW71722;

XX 11-JAN-1999 (first entry)

DE Arabidopsis leafy-cotyledon 1 (LEC1) polypeptide.

XX LEC1; leafy-cotyledon 1; embryo; transcription factor;
KW transgenic plant.

XX Arabidopsis thaliana ecotype Wassilewskija.

XX Key Location/Qualifiers

XX Binding-site 34..40 /note= "putative DNA binding site"

XX Domain 61..72 /note= "putative subunit interaction domain"

XX WO9637184-A1.

XX 27-AUG-1998.

XX 20-FEB-1998; 98WO-US02998.

XX 19-FEB-1998; 98US-0804534.

XX 21-FEB-1997; 97US-0804534.

XX (REGC) UNIV CALIFORNIA.

XX Fischer RL, Goldberg RB, Harada JT, Lotan T, Ohto M;

XX WPI: 1998-531499/45.

XX N-PSDB; AAV61022-24.

XX New isolated plant leafy cotyledon-1 gene - used to develop products
PT for, e.g. increasing storage protein content in plant tissues, or
PT producing fruit with small and non-viable seed
XX
XX

XX Claim 15; Page 31; 55pp; English.

XX LEC1 polypeptide is encoded by the leafy cotyledon-1 (LEC1) gene
CC (see AAV61022-24) of Arabidopsis thaliana. Full-length LEC1
CC polypeptide can act as a subunit of a protein capable of acting
CC as a transcription factor in plant cells. The LEC1 gene is
CC embryo-specific and can be used to modulate development (claimed) of
CC embryos or other organs in plants. Inhibiting expression can be
CC useful e.g. in weed control (by transferring an inhibitory sequence

CC to a weedy species and allowing it to be transmitted through sexual
CC crosses) or to produce fruit with small and non-viable seed.
CC Enhanced expression of LEC1 can be used to increase storage protein
CC content in plant tissues to improve nutritional value.

SO Sequence 208 AA;

Query Match 2.2%; Score 7; DB 19; Length 208;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 314 YVDPLTV 320
|||
105 YVDPLTV 111

RESULT 30
AAV54563

ID AAV54563 standard; Protein; 208 AA.

AC AAV54563;

DT 25-APR-2000 (first entry)

XX Amino acid sequence of an Arabidopsis leafy cotyledon 1 polypeptide.

XX Leafy cotyledon 1 gene; LEC1 gene; embryogenesis; cotyledon identity;
KW embryo development; transcription factor; seed development;
KW ecotopic development; embryonic plant tissue; weed control;
KW nutritional value; storage protein; cotyledon; seed;
KW reproductive tissue mass; fruit size; seed mass.

OS Arabidopsis thaliana.

PN W09967405-A2.

PD 29-DEC-1999.

PF 24-JUN-1999; 99WO-US14384.

PR 24-JUN-1998; 98US-0103478.

PR 17-NOV-1998; 98US-0193931.

PA (REGC) UNIV CALIFORNIA.

PI Harada JT, Lotan T, Ohto M, Goldberg RB, Fischer RL;

DR WPI; 2000-160588/14.

DR N-PSDB; AAZ45838.

PT New embryo-specific gene useful for producing transgenic plant

XX Claim 6; Page 53-54; 69pp; English.

CC The present sequence represents a leafy Cotyledon 1 (LEC1) polypeptide.
CC LEC1 genes are thought to play a central role in late embryogenesis, in
CC specifying cotyledon identity during embryo development. LEC1
CC polypeptides may act as transcription factors. LEC1 polynucleotides are
CC useful for modulating seed development and for inducing ecotopic
CC development of embryonic tissue in a plant. In both cases, the LEC1
CC polynucleotide is introduced into the plant through a sexual cross and
CC is co-expressed in an antisense orientation with a second heterologous
CC polynucleotide selected from AP2 and RAR2 genes of Arabidopsis. The
CC LEC1 polynucleotide is also useful for targeting expression in a seed,
CC and for preparing expression cassettes for suppressing or enhancing
CC endogenous LEC1 gene expression, which is useful in weed control or
CC for improving nutritional value of plant tissue respectively. LEC1
CC polypeptides and polynucleotides are especially used for increasing or
CC decreasing storage protein content in cotyledons or leaves. LEC1 also
CC increases reproductive tissue mass, e.g., increases fruit size, seed
CC mass, its protein or its oil.

XX Sequence 208 AA;

Query Match 2.2%; Score 7; DB 21; Length 208;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 314 YVDPLTV 320
|||
105 YVDPLTV 111

RESULT 31

ID AAE09828 standard; Protein; 208 AA.

AC AAE09828;

DT 29-NOV-2001 (first entry)

XX Leafy cotyledon 1 (LEC1) protein.

XX Leafy cotyledon 1; LEC1; transcription modulator; seed development;
KW embryonic characteristic; transgenic plant; storage protein content;
KW weed control; ectopic expression.

XX Unidentified.

XX Key Location/Qualifiers

FT Domain 28..117
/label=B-domain

XX W0200164022-A2.

XX 07-SEP-2001.

XX 21-FEB-2001; 2001WO-US05454.

XX 01-MAR-2000; 2000US-0516052.

PA (REGC) UNIV CALIFORNIA.

PI Harada JT, Lotan T, Ohto M, Goldberg RB, Fischer RL, Bui A;

PI Kwong R;

DR WPI; 2001-565462/63.

DR N-PSDB; AAD16914, AAD16915, AAD16916.

PT Novel leafy cotyledon 1 polynucleotide encoding leafy cotyledon

PT polypeptide for modulating transcription resulting in seed development

PT in plant comprises providing in an expression cassette linked to LEC1

XX gene promoter -

XX Claim 2; Page 53-54; 73pp; English.

CC The patent discloses novel leafy cotyledon 1 (LEC1) polypeptides and
CC polynucleotides encoding them. The invention also provides expression
CC cassettes comprising a promoter operably linked to a heterologous
CC sequence that encodes LEC-1 protein comprising a subsequence that
CC is similar to LEC-1 B domain. LEC-1 sequences are used for modulating
CC transcription resulting in the induction of embryonic characteristics
CC or seed development in a plant. Polynucleotide sequences from LEC1
CC genes are used to direct expression of desired heterologous genes in
CC embryos, to modulate the development of embryos or the characteristics
CC of an embryo on other organs (e.g., by enhancing expression of the gene
CC in a transgenic plant). Modulation of the expression of LEC-1 gene is
CC used to manipulate a number of useful traits, such as increasing or
CC decreasing storage protein content in cotyledons or leaves. Inhibition
CC of LEC1 gene expression is useful in weed control or to produce fruit
CC with small and non-viable seed. Enhanced expression of LEC1 genes is
CC useful to increase storage protein content in plant tissues. The LEC-1
CC gene promoters are useful for directing gene expression so that the
CC desired gene product is located in embryos or seeds. Expression of
CC LEC-1 genes in plant reproductive or vegetative tissue induce ectopic
CC expression of cells and tissues, together with another plant nucleic

PR	06-SEP-2000	2000US-02394637
PR	06-SEP-2000	2000US-02394637
PR	06-SEP-2000	2000US-02304338
PR	08-SEP-2000	2000US-02311442
PR	08-SEP-2000	2000US-02311442
PR	08-SEP-2000	2000US-02311443
PR	08-SEP-2000	2000US-02311443
PR	08-SEP-2000	2000US-02311414
PR	08-SEP-2000	2000US-02311414
PR	08-SEP-2000	2000US-02320801
PR	08-SEP-2000	2000US-02320801
PR	12-SEP-2000	2000US-02311668
PR	14-SEP-2000	2000US-0232197
PR	14-SEP-2000	2000US-0232197
PR	14-SEP-2000	2000US-02323998
PR	14-SEP-2000	2000US-02323998
PR	14-SEP-2000	2000US-0232401
PR	14-SEP-2000	2000US-0232401
PR	14-SEP-2000	2000US-02330664
PR	14-SEP-2000	2000US-02330664
PR	21-SEP-2000	2000US-02344223
PR	21-SEP-2000	2000US-02344223
PR	25-SEP-2000	2000US-02344977
PR	25-SEP-2000	2000US-02344977
PR	25-SEP-2000	2000US-02345498
PR	26-SEP-2000	2000US-02354884
PR	27-SEP-2000	2000US-02356334
PR	27-SEP-2000	2000US-02356334
PR	29-SEP-2000	2000US-0236127
PR	29-SEP-2000	2000US-0236127
PR	29-SEP-2000	2000US-02361687
PR	29-SEP-2000	2000US-02361687
PR	29-SEP-2000	2000US-02365699
PR	29-SEP-2000	2000US-02365699
PR	02-OCT-2000	2000US-0236802
PR	02-OCT-2000	2000US-0236802
PR	02-OCT-2000	2000US-0237037
PR	02-OCT-2000	2000US-0237037
PR	02-OCT-2000	2000US-02370338
PR	02-OCT-2000	2000US-02370338
PR	02-OCT-2000	2000US-02370409
PR	13-OCT-2000	2000US-02399335
PR	13-OCT-2000	2000US-02399337
PR	20-OCT-2000	2000US-0240960
PR	20-OCT-2000	2000US-02411221
PR	20-OCT-2000	2000US-0241185
PR	20-OCT-2000	2000US-02417867
PR	20-OCT-2000	2000US-02417867
PR	20-OCT-2000	2000US-0241808
PR	20-OCT-2000	2000US-0241809
PR	20-OCT-2000	2000US-0241826
PR	01-NOV-2000	2000US-0244617
PR	08-NOV-2000	2000US-0246474
PR	08-NOV-2000	2000US-0246475
PR	08-NOV-2000	2000US-0246476
PR	08-NOV-2000	2000US-0246477
PR	08-NOV-2000	2000US-0246513
PR	08-NOV-2000	2000US-0246517
PR	08-NOV-2000	2000US-0246528
PR	17-NOV-2000	2000US-0249207
PR	17-NOV-2000	2000US-0249207
PR	17-NOV-2000	2000US-0249209
PR	17-NOV-2000	2000US-0249210
PR	17-NOV-2000	2000US-0249211
PR	17-NOV-2000	2000US-0249212
PR	17-NOV-2000	2000US-0249213
PR	17-NOV-2000	2000US-0249214
PR	17-NOV-2000	2000US-0249215
PR	17-NOV-2000	2000US-0249216

PR 17-NOV-2000; 2000US-02492317.
PR 17-NOV-2000; 2000US-02492418.
FR 17-NOV-2000; 2000US-0249244.
FR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
F1 Rosen CA, Baraash SC, Ruben SM;
XX WPI; 2001-465572/50.
DR N-PSDB; AAS31221.
XX

PT Nucleic acid molecules encoding human secreted extracellular matrix
PT proteins, used in preventing, treating or ameliorating a disorder, e.g.
PT Alzheimer's and Parkinson's diseases and cancers -
XX

PS Claim 11; SEQ ID NO 300; 577pp; English.

XX
XX The invention relates to isolated nucleic acid molecules encoding
CC novel human secreted extracellular matrix proteins (SPs). The
CC polynucleotides and proteins are used to prevent, treat a medical
CC condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs,
CC chickens or sheep. For example, disorders associated with decreased
CC expression of SPs. The SP polynucleotide or a vector expressing them may
CC be administered to treat diseases by gene therapy. Antisense molecules
CC may be administered to down regulate expression of SPs by binding with
CC the cells own genes and preventing their expression. The polynucleotides
CC may also be used as DNA probes in diagnostic assays. The SPs may also be
CC used as antigens to produce antibodies and to identify modulators
CC (agonists and antagonists) of the SPs. The anti-(SP) antibodies and
CC antagonists may also be used to down regulate expression and activity of
CC SP and as diagnostic agents for detecting the presence of SPs in samples.
CC The disorders include for example: immune/autoimmune diseases (e.g. HIV
CC (human immunodeficiency virus) infections, anemia, rheumatoid arthritis
CC and multiple sclerosis), cancers and hyperproliferative disorders (e.g.
CC melanomas, neoplasms of the breast or liver, Sezary syndrome and
CC Gaucher's disease), neurological diseases (e.g. Alzheimer's disease,
CC Parkinson's disease), cardio-/cerebrovascular disorders (e.g. cardiac
CC arrest, tachycardia and angina), infections caused by bacteria, viruses
CC and fungi and ocular disorders (e.g. corneal infections). Other uses
CC include wound healing, maintenance of organs before transplantation,

Query Match 2.2%; Score 7; DB 22; Length 251;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 167 KPFTPP 173
|||||
Db 211 KPFTPP 217

RESULT 34
ABP47870
ID ABP47870 standard; Protein; 251 AA.
XX

AC ABP47870;
 XX 23-AUG-2002 (first entry)
 DT
 XX
 DE Human polypeptide SEQ ID NO 300.
 XX
 KW Human; noctropic; neuroprotective; cytosolic; dermatological; virucide;
 KW immunosuppressive; anti-inflammation; anti-HIV; antibacterial; vulnery;
 KW antiparkinsonian; antiscikling; antianemic; antithratic; cancer;
 KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
 KW antiallergic; antidiabetic; antifulcer; anticonvulsant; antifungal;
 KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;
 KW neurological disease; infection; nephrotropic; gene therapy; vaccine.
 XX
 OS Homo sapiens.
 XX
 XX US2002042386-A1.
 XX
 PD 11-APR-2002.
 XX
 PF 17-JAN-2001; 2001US-0764870.
 XX
 PR 31-JAN-2000; 2000US-179065P.
 PR 04-FEB-2000; 2000US-180528P.
 PR 28-JUN-2000; 2000US-214868P.
 PR 07-JUL-2000; 2000US-216647P.
 PR 07-JUL-2000; 2000US-216680P.
 PR 11-JUL-2000; 2000US-217487P.
 PR 14-JUL-2000; 2000US-218290P.
 PR 26-JUL-2000; 2000US-220963P.
 PR 14-AUG-2000; 2000US-220964P.
 PR 14-AUG-2000; 2000US-224518P.
 PR 14-AUG-2000; 2000US-224519P.
 PR 14-AUG-2000; 2000US-225267P.
 PR 14-AUG-2000; 2000US-225270P.
 PR 14-AUG-2000; 2000US-225477P.
 PR 14-AUG-2000; 2000US-225757P.
 PR 14-AUG-2000; 2000US-225758P.
 PR 22-AUG-2000; 2000US-226868P.
 PR 30-AUG-2000; 2000US-228924P.
 PR 01-SEP-2000; 2000US-229287P.
 PR 01-SEP-2000; 2000US-229343P.
 PR 01-SEP-2000; 2000US-229344P.
 PR 05-SEP-2000; 2000US-229345P.
 PR 05-SEP-2000; 2000US-229509P.
 PR 08-SEP-2000; 2000US-231413P.
 PR 21-SEP-2000; 2000US-234233P.
 PR 21-SEP-2000; 2000US-234274P.
 PR 27-SEP-2000; 2000US-234597P.
 PR 29-SEP-2000; 2000US-235834P.
 PR 29-SEP-2000; 2000US-236327P.
 PR 29-SEP-2000; 2000US-236367P.
 PR 29-SEP-2000; 2000US-236388P.
 PR 29-SEP-2000; 2000US-236396P.
 PR 29-SEP-2000; 2000US-236370P.
 PR 02-OCT-2000; 2000US-236802P.
 PR 02-OCT-2000; 2000US-237037P.
 PR 02-OCT-2000; 2000US-237038P.
 PR 02-OCT-2000; 2000US-237039P.
 PR 13-OCT-2000; 2000US-239935P.
 PR 20-OCT-2000; 2000US-240960P.
 PR 20-OCT-2000; 2000US-241785P.
 PR 20-OCT-2000; 2000US-241809P.
 PR 17-NOV-2000; 2000US-244617P.
 PR 17-NOV-2000; 2000US-249299P.
 PR 08-DEC-2000; 2000US-251856P.
 PR 08-DEC-2000; 2000US-251868P.
 PR 08-DEC-2000; 2000US-251869P.
 XX

PA (ROSE/) ROSEN C A.
 PA (RUBEN/) RUBEN S M.
 PA (BARA/) BARASH S C.
 XX
 PI Rosen CA, Ruben SM, Barash SC;
 XX
 DR WPI; 2002-470713/50.
 DR N-Psdb; ABQ66545.
 XX
 PT New nucleic acid encoding human proteins, useful for diagnosis,
 PT treatment and prevention of e.g. osteoporosis, also related
 PT polypeptides and antibodies -
 XX
 PS Claim 11; SEQ ID NO 300; 235pp + Sequence Listing; English.
 XX
 CC The invention relates to novel genes (ABQ66521-ABQ66785) and proteins
 CC (ABP47846-ABP48110) useful for preventing, treating or ameliorating
 CC medical conditions e.g. by protein or gene therapy. The genes are
 CC isolated from a range of human tissues disclosed in the specification.
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
 CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
 CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
 CC marrow, breast, gastrointestinal tract, liver, lung or urogenital;
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
 CC and parasitic infections.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from USPTO at seqdata.uspto.gov/sequence.html?docID=99909764870.
 XX
 SQ Sequence 251 AA;
 XX
 Query Match 2.2% Score 7; DB 23; Length 251;
 Best Local Similarity 100.0%; Pred. No. 2e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 167 KPTPRP 173
 Db 211 KPTPRP 217
 XX
 RESULT 35
 ID ABP30505 standard; Protein; 236 AA.
 XX
 AC ABP30505;
 XX
 DT 02-JUL-2002 (first entry)
 XX
 DE Streptococcus polypeptide SEQ ID NO 10186.
 XX
 KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
 KW group A streptococcus; Streptococcus pyogenes; antibacterial;
 KW antiinflammatory; infection; vaccine; meningitis; gene therapy.
 XX
 OS Streptococcus agalactiae.
 XX
 PN WO200234771-A2.
 PD 02-MAY-2002.
 XX
 PF 29-OCT-2001; 2001WO-GB04789.
 XX
 PR 27-OCT-2000; 2000GB-0026333.
 PR 24-NOV-2000; 2000GB-0028727.
 PR 07-MAR-2001; 2001GB-0005640.
 XX
 PA (CHIR-) CHIRON SPA.
 PA (GENO-) INST GENOMIC RES.

```

XX  Telford J, Massignani V, Margarit Ros YI, Grandi G, Fraser C,
PI  Tettein H;
XX  WPI; 2002-352536/38.
DR  N-PSDB; ABN71136.
XX
PT  New Streptococcus protein for the treatment or prevention of infection
PT  or disease caused by Streptococcus bacteria, such as meningitis, and
PT  for detecting a compound that binds to the protein -
PS  Claim 1; Page 4148; 4525pp; English.
XX
CC  The invention relates to a protein (ABP25413-ABP30895) from group B
CC  streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
CC  (Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in
CC  the specification. The proteins have antibacterial and antiinflammatory
CC  activity. (I), nucleic acids encoding (I), ABN6044-ABN71526 and
CC  antibodies that bind (I) are used in the manufacture of medicaments for
CC  the treatment or prevention of infection or disease caused by
CC  Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC  Nucleic acids encoding (I) are used to detect Streptococcus in a
CC  biological sample. (I) is used to determine whether a compound binds to
CC  (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC  used as a vaccine or diagnostic composition. The disease caused by
CC  Streptococcus that is prevented or treated may be meningitis. Nucleic
CC  acid encoding (I) may be used to recombinantly produce (I) and may be
CC  used in gene therapy. Antibodies to (I) are used for affinity
CC  chromatography, immunoassays, and distinguishing/identifying
CC  Streptococcus proteins.
XX
SQ  Sequence 298 AA;

Query Match 2.2%; Score 7; DB 23; Length 298;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 63 ITDSQGV 69
Db 24 ITDSQGV 30

RESULT 36
ABB08497
ID ABB08497 standard; Protein; 300 AA.
XX
AC ABB08497;
XX
DT 23-MAY-2002 (first entry)
XX
DE Envelope 2 protein of HCV.
XX
KW Envelope 2 protein; HCV.
XX
OS Unidentified.
XX
PN KR138597-B1.
XX
PD 30-APR-1998.
XX
PF 30-JUL-1994; 94KR-0018833.
XX
PR 30-JUL-1994; 94KR-0018833.
XX
PA (GLDS ) LG CHEM CO LTD.
XX
PI Ryu W, Yang J, Cho J;
XX
DR WPI; 2000-144176/13.
XX
DR N-PSDB; ABL41548.
XX
PT STABILISED RECOMBINANT ANIMAL CELL LINE EXPRESSING ENVELOPE 2 PROTEIN
PT OF HCV

```

```

XX  Disclosure; Page 7; 8pp; Korean.
XX
PS  This invention relates to stabilised recombinant animal cell line
CC  expressing envelope 2 protein of HCV. The present sequence
CC  represents the envelope 2 protein of HCV.
XX
SQ  Sequence 300 AA;

Query Match 2.2%; Score 7; DB 21; Length 300;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 136 TPSPVAV 142
Db 141 TPSPVAV 147

RESULT 37
AAR98460
ID AAR98460 standard; Protein; 304 AA.
XX
AC AAR98460;
XX
DT 18-DEC-1996 (first entry)
XX
DE Endonuclease LlaIIc.
XX
KW LlaIIA; LlaIIB; LlaIIC; m6a-methyltransferase; endonuclease; restriction;
KW modification system; phage resistance; milk; dairy product;
KW food additive.
XX
OS Lactococcus lactis SMQ-17.
XX
PN W09621017-A2.
XX
PD 11-JUL-1996.
XX
PF 29-DEC-1995; 95WO-NL00448.
XX
PR 19-APR-1995; 95US-0424641.
XX
PR 30-DEC-1994; 94US-0366480.
XX
PA (UNIL ) QUEST INT BV.
XX
PI Moineau S, Vandenberg PA, Vedamuthu ER, Walker SA.
XX
DR WPI; 1996-333992/33.
XX
DR N-PSDB; AAT30400.
XX
PT DNA encoding restriction and modification enzymes from L. lactis -
PT confers phage resistance to bacteria, for use in milk and as a food
PT additive
XX
PS Claim 2; Fig 4; 48pp; English.
XX
CC The present sequence is that of the LlaIIC gene product, an
CC endonuclease, part of a Lactococcus lactis restriction/modification (R/M)
CC system. The LlaII R/M system is similar to the MboI system which
CC recognises the sequence 5'-GATC-3' and cleaves it before the guanine. A
CC plasmid contg. the DNA (AAT30400) encoding the R/M system can confer
CC phage resistance to a bacterium, such as providing resistance to
CC phage-sensitive dairy cultures. The DNA and proteins can be used for
CC improving fermentation of a dairy prod. and as additives in food
CC processing.
XX
SQ Sequence 304 AA;

Query Match 2.2%; Score 7; DB 17; Length 304;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 80 TVSKIAQ 86

```


hybridisation assay; genetic mapping; gene expression control; promoter;
termination sequence.
Arabidopsis thaliana.
EP1033405-A2.
06-SEP-2000.
25-FEB-2000; 2000EP-0301439.
25-FEB-1999; 99US-011825.
05-MAR-1999; 99US-0123180.
09-MAR-1999; 99US-0123548.
23-MAR-1999; 99US-0125788.
25-MAR-1999; 99US-0126264.
29-MAR-1999; 99US-0126785.
01-APR-1999; 99US-0127462.
06-APR-1999; 99US-0128234.
16-APR-1999; 99US-0128714.
19-APR-1999; 99US-0130077.
21-APR-1999; 99US-0130449.
23-APR-1999; 99US-0130510.
28-APR-1999; 99US-0130891.
30-APR-1999; 99US-0131449.
04-MAY-1999; 99US-0132407.
05-MAY-1999; 99US-0132484.
06-MAY-1999; 99US-0132485.
07-MAY-1999; 99US-0132486.
11-MAY-1999; 99US-0132863.
14-MAY-1999; 99US-0134218.
14-MAY-1999; 99US-0134219.
18-MAY-1999; 99US-0134370.
19-MAY-1999; 99US-0134766.
20-MAY-1999; 99US-0134941.
21-MAY-1999; 99US-0135124.
24-MAY-1999; 99US-0135353.
25-MAY-1999; 99US-0135629.
27-MAY-1999; 99US-0136021.
28-MAY-1999; 99US-0136392.
01-JUN-1999; 99US-0136782.
03-JUN-1999; 99US-0137528.
04-JUN-1999; 99US-0137502.
07-JUN-1999; 99US-0137724.
08-JUN-1999; 99US-0138094.
10-JUN-1999; 99US-0138540.
14-JUN-1999; 99US-0138847.
16-JUN-1999; 99US-0139119.
16-JUN-1999; 99US-0139452.
17-JUN-1999; 99US-0139453.
18-JUN-1999; 99US-0139454.
18-JUN-1999; 99US-0139455.
18-JUN-1999; 99US-0139456.
18-JUN-1999; 99US-0139457.
18-JUN-1999; 99US-0139458.
18-JUN-1999; 99US-0139459.
18-JUN-1999; 99US-0139460.
18-JUN-1999; 99US-0139461.
18-JUN-1999; 99US-0139462.
18-JUN-1999; 99US-0139463.
18-JUN-1999; 99US-0139750.
21-JUN-1999; 99US-0139817.
22-JUN-1999; 99US-0139899.
23-JUN-1999; 99US-0140353.
23-JUN-1999; 99US-0140354.
24-JUN-1999; 99US-0140695.
28-JUN-1999; 99US-0140823.
29-JUN-1999; 99US-0140991.
30-JUN-1999; 99US-0141287.
01-JUL-1999; 99US-0141842.
01-JUL-1999; 99US-0142154.
02-JUL-1999; 99US-0142055.
06-JUL-1999; 99US-0142390.
08-JUL-1999; 99US-0142803.
09-JUL-1999; 99US-0142920.
12-JUL-1999; 99US-0142977.
13-JUL-1999; 99US-0143542.
14-JUL-1999; 99US-0143624.
15-JUL-1999; 99US-0144005.
16-JUL-1999; 99US-0144085.
16-JUL-1999; 99US-0144086.
19-JUL-1999; 99US-0144325.
19-JUL-1999; 99US-0144331.
19-JUL-1999; 99US-0144332.
19-JUL-1999; 99US-0144333.
19-JUL-1999; 99US-0144334.
19-JUL-1999; 99US-0144335.
20-JUL-1999; 99US-0144352.
20-JUL-1999; 99US-0144632.
20-JUL-1999; 99US-0144884.
21-JUL-1999; 99US-0144814.
21-JUL-1999; 99US-0145086.
21-JUL-1999; 99US-0145088.
22-JUL-1999; 99US-0145085.
22-JUL-1999; 99US-0145087.
22-JUL-1999; 99US-0145089.
22-JUL-1999; 99US-0145192.
23-JUL-1999; 99US-0145145.
23-JUL-1999; 99US-0145218.
23-JUL-1999; 99US-0145224.
26-JUL-1999; 99US-0145276.
27-JUL-1999; 99US-0145913.
27-JUL-1999; 99US-0145918.
27-JUL-1999; 99US-0145919.
28-JUL-1999; 99US-0145951.
28-AUG-1999; 99US-0146386.
02-AUG-1999; 99US-0146388.
02-AUG-1999; 99US-0146389.
03-AUG-1999; 99US-0147038.
04-AUG-1999; 99US-0147204.
04-AUG-1999; 99US-0147302.
05-AUG-1999; 99US-0147192.
05-AUG-1999; 99US-0147260.
06-AUG-1999; 99US-0147303.
06-AUG-1999; 99US-0147416.
09-AUG-1999; 99US-0147493.
09-AUG-1999; 99US-0147935.
10-AUG-1999; 99US-0148171.
11-AUG-1999; 99US-0148319.
12-AUG-1999; 99US-0148341.
13-AUG-1999; 99US-0148565.
13-AUG-1999; 99US-0148664.
16-AUG-1999; 99US-0149368.
17-AUG-1999; 99US-0149175.
18-AUG-1999; 99US-0149426.
20-AUG-1999; 99US-0149722.
20-AUG-1999; 99US-0149723.
20-AUG-1999; 99US-0149929.
23-AUG-1999; 99US-0149902.
23-AUG-1999; 99US-0149930.
23-AUG-1999; 99US-0149936.
25-AUG-1999; 99US-0150566.
26-AUG-1999; 99US-0150884.
27-AUG-1999; 99US-0151065.
27-AUG-1999; 99US-0151066.
27-AUG-1999; 99US-0151067.
30-AUG-1999; 99US-0151303.
31-AUG-1999; 99US-0151348.
01-SEP-1999; 99US-0151930.

PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match

2.2%; Score 7; DB 21; Length 337;

Best Local Similarity 100.0%; Pred.No.2.6e+02; Mismatches 0; Indels 0; Gaps 0;

Matches 7; Conservative

Oy 122 VRERSIS 128
Db 129 VRERSIS 135

Search completed: July 8, 2003, 11:18:50
Job time : 95 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: July 8, 2003, 11:17:18 / Search time 27 Seconds
(without alignments)
350.896 Million cell updates/sec

Title: US-10-018-706-2
Perfect score: 322
Sequence: 1 MVTAINSONQKPIKRLGL.....LFEFRISRNQGVYDPLTLK 322

Scoring table: OLIGO
Gapop 60.0 , Gapept 60.0

Searched: 262574 seqs, 29422922 residues

Word size : 0
Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database :

Issued Patents AA:*
1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/6C_COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/backfill1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8	2.5	557	US-09-027-064-2	Sequence 2, Appl
2	8	2.5	557	US-09-271-815-2	Sequence 2, Appl
3	7	2.2	13	US-08-258-152-9	Sequence 9, Appl
4	7	2.2	13	US-08-076-299A-9	Sequence 9, Appl
5	7	2.2	13	US-08-438-582-9	Sequence 9, Appl
6	7	2.2	13	US-09-266-596-9	Sequence 9, Appl
7	7	2.2	13	US-08-479-737-44	Sequence 9, Appl
8	7	2.2	13	US-08-475-442A-44	Sequence 44, Appl
9	7	2.2	56	US-09-029-424-15	Sequence 15, Appl
10	7	2.2	90	US-09-103-478-19	Sequence 19, Appl
11	7	2.2	90	US-09-193-931C-19	Sequence 19, Appl
12	7	2.2	208	US-09-103-478-2	Sequence 2, Appl
13	7	2.2	208	US-09-193-931C-2	Sequence 2, Appl
14	7	2.2	304	US-08-424-641B-4	Sequence 4, Appl
15	7	2.2	304	US-08-820-980-4	Sequence 4, Appl
16	7	2.2	304	US-08-826-439-4	Sequence 4, Appl
17	7	2.2	304	US-08-913-159-8	Sequence 8, Appl
18	7	2.2	344	US-09-147-236-11	Sequence 11, Appl
19	7	2.2	350	US-08-909-828-1	Sequence 1, Appl
20	7	2.2	350	US-08-909-828-2	Sequence 2, Appl
21	7	2.2	356	US-09-232-191-4	Sequence 4, Appl
22	7	2.2	356	US-09-232-200-4	Sequence 4, Appl
23	7	2.2	356	US-09-232-197-4	Sequence 4, Appl
24	7	2.2	356	US-09-232-201-4	Sequence 4, Appl
25	7	2.2	501	US-08-149-105-15	Sequence 15, Appl
26	7	2.2	501	US-08-317-847-15	Sequence 15, Appl
27	7	2.2	505	US-09-382-256-8	Sequence 8, Appl

28	7	2.2	505	US-09-382-256-16	Sequence 16, Appl
29	7	2.2	505	US-09-395-115-8	Sequence 8, Appl
30	7	2.2	505	US-09-395-115-16	Sequence 16, Appl
31	7	2.2	505	US-08-123-934A-8	Sequence 8, Appl
32	7	2.2	505	US-08-436-265-8	Sequence 8, Appl
33	7	2.2	505	US-08-436-265-16	Sequence 16, Appl
34	7	2.2	505	US-09-679-187-16	Sequence 8, Appl
35	7	2.2	505	US-09-679-187-16	Sequence 16, Appl
36	7	2.2	505	PCT-US94-10080-8	Sequence 8, Appl
37	7	2.2	543	US-08-469-412A-7	Sequence 7, Appl
38	7	2.2	543	US-09-021-715-7	Sequence 7, Appl
39	7	2.2	623	US-09-232-200-9	Sequence 9, Appl
40	7	2.2	623	US-09-232-200-98	Sequence 98, Appl
41	7	2.2	623	US-09-232-197-89	Sequence 89, Appl
42	7	2.2	623	US-09-232-197-88	Sequence 88, Appl
43	7	2.2	623	US-09-232-201-89	Sequence 89, Appl
44	7	2.2	623	US-09-232-201-98	Sequence 98, Appl
45	7	2.2	843	US-09-235-451-25	Sequence 25, Appl
46	6	1.9	8	US-08-803-899-18	Sequence 18, Appl
47	6	1.9	8	US-09-029-424-19	Sequence 19, Appl
48	6	1.9	9	US-08-146-145-3	Sequence 3, Appl
49	6	1.9	9	US-08-481-985B-130	Sequence 130, App
50	6	1.9	10	US-08-476-509B-31	Sequence 31, Appl
51	6	1.9	11	US-08-803-899-20	Sequence 20, Appl
52	6	1.9	12	US-08-803-899-19	Sequence 19, Appl
53	6	1.9	13	US-08-803-899-12	Sequence 12, Appl
54	6	1.9	13	US-08-803-899-17	Sequence 17, Appl
55	6	1.9	14	US-08-061-092A-11	Sequence 11, Appl
56	6	1.9	14	US-08-481-003-31	Sequence 31, Appl
57	6	1.9	14	US-08-256-156A-1	Sequence 1, Appl
58	6	1.9	14	US-08-373-190-6	Sequence 6, Appl
59	6	1.9	14	US-08-224-591-8	Sequence 8, Appl
60	6	1.9	14	US-08-392-138A-2	Sequence 2, Appl
61	6	1.9	14	US-08-438-190A-6	Sequence 6, Appl
62	6	1.9	14	US-08-926-789-8	Sequence 8, Appl
63	6	1.9	14	US-08-818-253-12	Sequence 12, Appl
64	6	1.9	14	US-08-818-253-44	Sequence 44, Appl
65	6	1.9	14	US-08-350-215-6	Sequence 6, Appl
66	6	1.9	14	US-09-166-750-2	Sequence 2, Appl
67	6	1.9	14	US-09-166-093-2	Sequence 2, Appl
68	6	1.9	14	US-09-287-145A-6	Sequence 6, Appl
69	6	1.9	14	US-08-485-598-31	Sequence 31, Appl
70	6	1.9	14	US-08-454-098-23	Sequence 23, Appl
71	6	1.9	14	US-09-172-019-2	Sequence 2, Appl
72	6	1.9	14	US-09-166-094-2	Sequence 2, Appl
73	6	1.9	14	US-08-501-253A-10	Sequence 10, Appl
74	6	1.9	14	US-08-818-252-42	Sequence 42, Appl
75	6	1.9	14	US-08-818-252-44	Sequence 44, Appl
76	6	1.9	14	US-09-362-805-4	Sequence 4, Appl
77	6	1.9	14	US-09-362-805-6	Sequence 6, Appl
78	6	1.9	14	US-09-173-190-6	Sequence 6, Appl
79	6	1.9	14	US-09-173-190-4	Sequence 4, Appl
80	6	1.9	14	US-09-069-821-7	Sequence 7, Appl
81	6	1.9	14	US-09-556-111-6	Sequence 6, Appl
82	6	1.9	14	US-09-316-919-58	Sequence 58, Appl
83	6	1.9	14	US-09-316-919-60	Sequence 60, Appl
84	6	1.9	14	PCT-US93-11138-8	Sequence 8, Appl
85	6	1.9	14	PCT-US96-01600-23	Sequence 23, Appl
86	6	1.9	15	US-08-687-956A-3	Sequence 3, Appl
87	6	1.9	15	US-08-687-956A-6	Sequence 6, Appl
88	6	1.9	18	US-08-224-591-1	Sequence 1, Appl
89	6	1.9	18	US-08-224-591-9	Sequence 9, Appl
90	6	1.9	18	US-08-224-591-10	Sequence 10, Appl
91	6	1.9	18	US-08-352-338A-3	Sequence 3, Appl
92	6	1.9	18	US-08-819-033-3	Sequence 3, Appl
93	6	1.9	18	US-08-926-789-1	Sequence 1, Appl
94	6	1.9	18	US-08-926-789-9	Sequence 9, Appl
95	6	1.9	18	US-08-926-789-10	Sequence 10, Appl
96	6	1.9	18	US-08-818-253-43	Sequence 43, Appl
97	6	1.9	18	US-08-818-253-45	Sequence 45, Appl
98	6	1.9	18	US-09-166-750-3	Sequence 3, Appl
99	6	1.9	18	US-09-166-093-3	Sequence 3, Appl
100	6	1.9	18	US-09-172-019-3	Sequence 3, Appl

977 5 1.6 80 1 US-08-137-800-41 Sequence 41, App1
978 5 1.6 80 1 US-08-477-383-41 Sequence 41, App1
979 5 1.6 80 1 US-08-487-174-41 Sequence 41, App1
980 5 1.6 80 2 US-08-480-750-41 Sequence 85, App1
981 5 1.6 80 2 US-08-553-501A-85 Sequence 104, App1
982 5 1.6 80 2 US-08-765-783A-104 Sequence 10, App1
983 5 1.6 80 3 US-08-554-840-10 Sequence 12, App1
984 5 1.6 80 3 US-08-554-840-12 Sequence 13, App1
985 5 1.6 80 3 US-08-554-840-13 Sequence 15, App1
986 5 1.6 80 3 US-08-554-840-15 Sequence 5, App1
987 5 1.6 80 3 US-08-894-017-5 Sequence 85, App1
988 5 1.6 80 3 US-09-205-231-85 Sequence 104, App1
989 5 1.6 80 4 US-09-416-557-104 Sequence 117, App1
990 5 1.6 80 4 US-08-818-112-117 Sequence 112, App1
991 5 1.6 80 4 US-08-818-112-117 Sequence 117, App1
992 5 1.6 80 4 US-09-056-556-117 Sequence 10, App1
993 5 1.6 80 4 US-08-925-339-12 Sequence 12, App1
994 5 1.6 80 4 US-08-925-339-13 Sequence 13, App1
995 5 1.6 80 4 US-08-925-339-15 Sequence 15, App1
996 5 1.6 80 4 US-09-072-596-112 Sequence 112, App1
997 5 1.6 81 3 US-09-554-840-11 Sequence 211, App1
998 5 1.6 81 4 US-09-554-805-211 Sequence 450, App1
999 5 1.6 81 4 US-08-856-207A-450
1000 5 1.6 81 4

ALIGNMENTS

RESULT 1
US-09-027-064-2
Sequence 2, Application US/09027064
Patent No. 6133006
GENERAL INFORMATION:
APPLICANT: SHABON, USMAN
APPLICANT: BERGSMAN, DEREK
TITLE OF INVENTION: A YAK-1 RELATED SERINE/THREONINE
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: RATNER & PRESTIA
STREET: P.O. BOX 980
CITY: VALLEY FORGE
STATE: PA
COUNTRY: USA
ZIP: 19482
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/027,064
FILING DATE: 20-FEB-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/053,924
FILING DATE: 28-JUL-1997
ATTORNEY/AGENT INFORMATION:
NAME: PRESTIA, PAUL F
REGISTRATION NUMBER: 23,031
REFERENCE/DOCKET NUMBER: GH-70172
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-407-0700
TELEFAX: 610-407-0701
TELEX: 846169
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 557 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein

US-09-027-064-2
Query Match 2.5%; Score 8; DB 4; Length 557;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 179 APVAPPT 186
DB 450 APVAPPT 457

RESULT 2
US-09-271-815-2
Sequence 2, Application US/09271815
Patent No. 6297036
GENERAL INFORMATION:
APPLICANT: SHABON, USMAN
APPLICANT: BERGSMAN, DEREK
TITLE OF INVENTION: A YAK-1 RELATED SERINE/THREONINE PROTEIN KINASE-HTLAR33
FILE REFERENCE: GH-70172-1
CURRENT APPLICATION NUMBER: US/09/271,815
CURRENT FILING DATE: 1999-03-18
EARLIER APPLICATION NUMBER: 09/027,064
EARLIER FILING DATE: 1998-02-20
EARLIER APPLICATION NUMBER: 60/053,924
EARLIER FILING DATE: 1997-07-28
NUMBER OF SEQ ID NOS: 6
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 557
TYPE: PRT
ORGANISM: Homo sapiens
US-09-271-815-2

Query Match 2.5%; Score 8; DB 4; Length 557;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 179 APVAPPT 186
DB 450 APVAPPT 457

RESULT 3
US-08-258-152-9
Sequence 9, Application US/08258152
Patent No. 5686279
GENERAL INFORMATION:
APPLICANT: FINER, MITCHELL H.
APPLICANT: ROBERTS, MARGO R.
APPLICANT: DULL, THOMAS J.
APPLICANT: ZSEBO, KRISTINA M.
APPLICANT: QIN, LU
TITLE OF INVENTION: METHOD FOR PRODUCTION OF HIGH TITER
TITLE OF INVENTION: VIRUS AND HIGH EFFICIENCY RETROVIRAL MEDIATED TRANSDUCTION
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: CELL GENESYS, INC.
STREET: 322 LAKESIDE DRIVE
CITY: FOSTER CITY
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94404
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/258,152
FILING DATE: 10-JUN-1994
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/076,299
FILING DATE: 11-JUN-1993
ATTORNEY/AGENT INFORMATION:
NAME: KRUPEN, KAREN I.
REGISTRATION NUMBER: 34,647
REFERENCE/DOCKET NUMBER: CELL 13.1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-358-9600 X131
TELEFAX: 415-349-7392
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 13 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-258-152-9

Query Match 2.2%; Score 7; DB 1; Length 13;
Best Local Similarity 100.0%; Pred. No. 5.6;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 STSGSGS 47
|||||
Db 2 STSGSGS 8

RESULT 4
US-08-076-299A-9
Sequence 9, Application US/08076299A
Patent No. 5834256
GENERAL INFORMATION:
APPLICANT: FINER, MITCHELL H.
APPLICANT: ROBERTS, MARGO R.
APPLICANT: DULL, THOMAS J.
APPLICANT: ZSEBO, KRISZTINA M.
APPLICANT: QIN, LU
TITLE OF INVENTION: METHOD FOR PRODUCTION OF HIGH TITER
TITLE OF INVENTION: VIRUS AND HIGH EFFICIENCY RETROVIRAL MEDIATED TRANSDUCTION
NUMBER OF INVENTION: OF MAMMALIAN CELLS
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESSEE: CELL GENESYS, INC.
STREET: 322 LAKESIDE DRIVE
CITY: FOSTER CITY
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/076,299A
FILING DATE: 11-JUN-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: KRUPEN, KAREN I.
REGISTRATION NUMBER: 34,647
REFERENCE/DOCKET NUMBER: CELL 13.0
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-358-9600 X131
TELEFAX: 415-349-7392
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 13 amino acids
TYPE: amino acid
STRANDEDNESS: single

TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-076-299A-9

Query Match 2.2%; Score 7; DB 2; Length 13;
Best Local Similarity 100.0%; Pred. No. 5.6;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 STSGSGS 47
|||||
Db 2 STSGSGS 8

RESULT 5
US-08-438-582-9
Sequence 9, Application US/08438582
Patent No. 5858740
GENERAL INFORMATION:
APPLICANT: FINER, MITCHELL H.
APPLICANT: ROBERTS, MARGO R.
APPLICANT: DULL, THOMAS J.
APPLICANT: ZSEBO, KRISZTINA M.
APPLICANT: QIN, LU
TITLE OF INVENTION: METHOD FOR PRODUCTION OF HIGH TITER
TITLE OF INVENTION: VIRUS AND HIGH EFFICIENCY RETROVIRAL MEDIATED TRANSDUCTION
NUMBER OF INVENTION: OF MAMMALIAN CELLS
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: CELL GENESYS, INC.
STREET: 322 LAKESIDE DRIVE
CITY: FOSTER CITY
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/438,582
FILING DATE: 10-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/258,152
FILING DATE: 10-JUN-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/076,299
FILING DATE: 11-JUN-93
ATTORNEY/AGENT INFORMATION:
NAME: KRUPEN, KAREN I.
REGISTRATION NUMBER: 34,647
REFERENCE/DOCKET NUMBER: CELL 13.2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-358-9600 X131
TELEFAX: 415-349-7392
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 13 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-438-582-9

Query Match 2.2%; Score 7; DB 2; Length 13;
Best Local Similarity 100.0%; Pred. No. 5.6;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 STSGSGS 47
|||||
Db 2 STSGSGS 8

RESULT 6
US-09-266-596-9
Sequence 9, Application US/09266596
Patent No. 6218187
GENERAL INFORMATION:
APPLICANT: FINER, MITCHELL H.
APPLICANT: DULL, THOMAS J.
APPLICANT: ZSEBO, KRISTINA M.
APPLICANT: COOKE, KEEGAN
APPLICANT: FARSON, DEBORAH A.
TITLE OF INVENTION: METHOD FOR PRODUCTION OF HIGH TITER
TITLE OF INVENTION: VIRUS AND HIGH EFFICIENCY RETROVIRAL MEDIATED TRANSDUCTION
NUMBER OF SEQUENCES: 48
CORRESPONDENCE ADDRESS:
ADDRESSEE: CELL GENESYS, INC.
STREET: 322 LAKESIDE DRIVE
CITY: FOSTER CITY
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94404
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/266,596
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/517,488
FILING DATE: 21-AUG-1995
APPLICATION NUMBER: US 08/258,152
FILING DATE: 10-JUN-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/076,299
FILING DATE: 11-JUN-1993
ATTORNEY/AGENT INFORMATION:
NAME: KRUPEN, KAREN I.
REGISTRATION NUMBER: 34,647
REFERENCE/DOCKET NUMBER: CELL 13.3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-358-9600 X131
TELEFAX: 415-349-7392
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 13 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-266-596-9

Query Match 2.2%; Score 7; DB 4; Length 13;
Best Local Similarity 100.0%; Pred. No. 5.6;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 41 STSGSGS 47
Db 2 STSGSGS 8

RESULT 7
US-08-479-737-44
Sequence 44, Application US/08479737
Patent No. 6319494
GENERAL INFORMATION:
APPLICANT: CAPON, DANIEL J
APPLICANT: WEISS, ARTHUR
APPLICANT: IRVING, BRIAN A
APPLICANT: ROBERTS, MARGO R

US-09-266-596-9
Sequence 9, Application US/09266596
Patent No. 6218187
GENERAL INFORMATION:
APPLICANT: FINER, MITCHELL H.
APPLICANT: DULL, THOMAS J.
APPLICANT: ZSEBO, KRISTINA M.
APPLICANT: COOKE, KEEGAN
APPLICANT: FARSON, DEBORAH A.
TITLE OF INVENTION: METHOD FOR PRODUCTION OF HIGH TITER
TITLE OF INVENTION: VIRUS AND HIGH EFFICIENCY RETROVIRAL MEDIATED TRANSDUCTION
NUMBER OF SEQUENCES: 48
CORRESPONDENCE ADDRESS:
ADDRESSEE: CELL GENESYS, INC.
STREET: 322 LAKESIDE DRIVE
CITY: FOSTER CITY
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94404
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/479,737
FILING DATE: 07-JUN-1995
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/238,405
FILING DATE: 05-MAY-1994
ATTORNEY/AGENT INFORMATION:
NAME: Mandel, Saralynn
REGISTRATION NUMBER: 31,853
REFERENCE/DOCKET NUMBER: Cell 5.3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 358-9600
TELEFAX: (415) 358-0803
INFORMATION FOR SEQ ID NO: 44:
SEQUENCE CHARACTERISTICS:
LENGTH: 13 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 44:
US-08-479-737-44

Query Match 2.2%; Score 7; DB 4; Length 13;
Best Local Similarity 100.0%; Pred. No. 5.6;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 41 STSGSGS 47
Db 2 STSGSGS 8

RESULT 8
US-08-475-442A-44
Sequence 44, Application US/08475442A
Patent No. 6407221
GENERAL INFORMATION:
APPLICANT: CAPON, DANIEL J
APPLICANT: WEISS, ARTHUR
APPLICANT: IRVING, BRIAN A
APPLICANT: ROBERTS, MARGO R
APPLICANT: ZSEBO, KRISTINA
TITLE OF INVENTION: CHIMERIC CHAINS FOR
TITLE OF INVENTION: RECEPTOR-ASSOCIATED SIGNAL TRANSDUCTION PATHWAYS
NUMBER OF SEQUENCES: 51
CORRESPONDENCE ADDRESS:
ADDRESSEE: CELL GENESYS, INC.
STREET: 322 LAKESIDE DRIVE
CITY: FOSTER CITY
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94404
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/475,442A
FILING DATE: 06-JUN-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/238,405
FILING DATE: 05-MAY-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/988,194
FILING DATE: 09-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/627,643
FILING DATE: 14-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US91/09431
FILING DATE: 12-DEC-1991
ATTORNEY/AGENT INFORMATION:
NAME: KRUPEN, KAREN I
REGISTRATION NUMBER: 34,647
REFERENCE/DOCKET NUMBER: CELLS.5
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415)358-9600x131
TELEFAX: (415)349-7392
INFORMATION FOR SEQ ID NO: 44:
SEQUENCE CHARACTERISTICS:
LENGTH: 13 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-475-442A-44

Query Match 2.2%; Score 7; DB 4; Length 13;
Best Local Similarity 100.0%; Pred. No. 5.6;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 STSGSGS 47
DB 2 STSGSGS 8

RESULT 9
US-09-029-424-15
Sequence 15, Application US/09029424A
Patent No. 6030795
GENERAL INFORMATION:
APPLICANT: Saitoh, Masao
APPLICANT: Miyazono, Kohel
APPLICANT: Ichijo, Hidenori
TITLE OF INVENTION: NUCLEIC ACID MOLECULE ENCODING TGF RECEPTOR HAVING MODIFIED
TITLE OF INVENTION: GROWTH INHIBITION, AND ITS USE
FILE REFERENCE: L0461/7027
CURRENT APPLICATION NUMBER: US/09/029,424A
CURRENT FILING DATE: 1998-04-28
EARLIER APPLICATION NUMBER: PCT/GB96/02179
EARLIER FILING DATE: 1996-09-04
NUMBER OF SEQ ID NOS: 19
SEQ ID NO 15
LENGTH: 56
TYPE: PRT
ORGANISM: Homo sapiens
US-09-029-424-15

Query Match 2.2%; Score 7; DB 3; Length 56;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 STSGSGS 47
DB 35 STSGSGS 41

RESULT 10
US-09-103-478-19
Sequence 19, Application US/09103478
Patent No. 6235975
GENERAL INFORMATION:
APPLICANT: Harada, John
APPLICANT: Lotan, Tamar
APPLICANT: Ohto, Masa-aki
APPLICANT: Goldberg, Robert B.
APPLICANT: Fischer, Robert L.
TITLE OF INVENTION: LEAFY COTYLEDON1 Genes and Their Uses
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/103,478
FILING DATE: 24-JUN-1998
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/026,221
FILING DATE: 19-FEB-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/804,534
FILING DATE: 21-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: Binhorn, Gregory P.
REGISTRATION NUMBER: 38,440
REFERENCE/DOCKET NUMBER: 023070-077611US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 90 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-103-478-19

Query Match 2.2%; Score 7; DB 4; Length 90;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 314 YVDPLTV 320
DB 78 YVDPLTV 84

RESULT 11
US-09-193-931C-19
Sequence 19, Application US/09193931C
Patent No. 6320102
GENERAL INFORMATION:
APPLICANT: Harada, John
APPLICANT: Lotan, Tamar
APPLICANT: Ohto, Masa-aki
APPLICANT: Goldberg, Robert B.
APPLICANT: Fischer, Robert L.
TITLE OF INVENTION: LEAFY COTYLEDON1 Genes and Their Uses
FILE REFERENCE: 023070-077620
CURRENT APPLICATION NUMBER: US/09/193,931C

CURRENT FILING DATE: 1998-11-17
PRIOR APPLICATION NUMBER: US 09/103,478
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: US 09/026,221
PRIOR FILING DATE: 1998-02-19
PRIOR APPLICATION NUMBER: US 08/804,534
PRIOR FILING DATE: 1997-02-21
NUMBER OF SEQ ID NOS: 29
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 19
LENGTH: 90
TYPE: PRT
ORGANISM: Arabidopsis thaliana
FEATURE:
NAME/KEY: DOMAIN
LOCATION: (1)..(90)
OTHER INFORMATION: LECT1 HAP3 subunit of CCAAT box-binding factor
OTHER INFORMATION: (CBF) protein B domain homolog
US-09-193-931C-19

Query Match 2.2%; Score 7; DB 4; Length 90;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 314 YVDPLTV 320
DB 78 YVDPLTV 84

RESULT 12
US-09-103-478-2
Sequence 2, Application US/09103478
Patent No. 6235975
GENERAL INFORMATION:
APPLICANT: Harada, John
APPLICANT: Lotan, Tamara
APPLICANT: Ohto, Masa-aki
APPLICANT: Goldberger, Robert B.
APPLICANT: Fischer, Robert L.
TITLE OF INVENTION: LEAFY COTYLEDON1 Genes and Their Uses
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/103,478
FILING DATE: 24-JUN-1998
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/026,221
FILING DATE: 19-FEB-1998
PRIOR APPLICATION DATA: US 08/804,534
FILING DATE: 21-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: Einhorn, Gregory P.
REGISTRATION NUMBER: 38,440
REFERENCE/DOCKET NUMBER: 023070-077611US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 208 amino acids

TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-103-478-2

Query Match 2.2%; Score 7; DB 4; Length 208;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 314 YVDPLTV 320
DB 105 YVDPLTV 111

RESULT 13
US-09-193-931C-2
Sequence 2, Application US/09193931C
Patent No. 6320102
GENERAL INFORMATION:
APPLICANT: Harada, John
APPLICANT: Lotan, Tamara
APPLICANT: Ohto, Masa-aki
APPLICANT: Goldberger, Robert B.
APPLICANT: Fischer, Robert L.
TITLE OF INVENTION: The Regenes of the University of California
FILE REFERENCE: 023070-077620
CURRENT APPLICATION NUMBER: US/09/193,931C
CURRENT FILING DATE: 1998-11-17
PRIOR APPLICATION NUMBER: US 09/103,478
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: US 09/026,221
PRIOR FILING DATE: 1998-02-19
PRIOR APPLICATION NUMBER: US 08/804,534
PRIOR FILING DATE: 1997-02-21
NUMBER OF SEQ ID NOS: 29
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 2
LENGTH: 208
TYPE: PRT
ORGANISM: Arabidopsis thaliana
US-09-193-931C-2

Query Match 2.2%; Score 7; DB 4; Length 208;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 314 YVDPLTV 320
DB 105 YVDPLTV 111

RESULT 14
US-08-424-641B-4
Sequence 4, Application US/08424641B
Patent No. 5824523
GENERAL INFORMATION:
APPLICANT: Sylvain Moineau, Shirley A.
APPLICANT: Walker, Ebenezer R. Vedamuthu,
APPLICANT: and Peter A. Vandenberg
TITLE OF INVENTION: Isolated DNA Encoding
TITLE OF INVENTION: Enzyme For Phage
TITLE OF INVENTION: Resistance
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ian C. McLeod
STREET: 2190 Commons Parkway
CITY: Okemos
STATE: Michigan
COUNTRY: USA
ZIP: 48864
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 360 Kb

MEDIUM TYPE: storage
COMPUTER: Acer
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/424,641B
FILING DATE: April 19, 1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/366,480
FILING DATE: December 30, 1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Ian C. McLeod
REGISTRATION NUMBER: 20,931
REFERENCE/DOCKET NUMBER: MT 4.1-151
TELECOMMUNICATION INFORMATION:
TELEPHONE: (517) 347-4100
TELEFAX: (517) 347-4103
TELEX: No. 5824523e
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 304 Amino Acids
TYPE: Amino Acid
STRANDEDNESS: Single
TOPOLOGY: Linear
MOLECULE TYPE: Peptide
US-08-424-641B-4

Query Match 2.2%; Score 7; DB 2; Length 304;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 80 TVSKIAQ 86
Db 171 TVSKIAQ 177

RESULT 15
US-08-820-980-4
Sequence 4, Application US/08820980
Patent No. 5925388
GENERAL INFORMATION:
APPLICANT: Sylvain Moineau, Shirley A.
APPLICANT: Walker, Ebenezer R. Vedamuthu,
APPLICANT: and Peter A. Vandenberg
TITLE OF INVENTION: Isolated DNA Encoding
TITLE OF INVENTION: Enzyme For Phage
TITLE OF INVENTION: Resistance
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ian C. McLeod
STREET: 2190 Commons Parkway
CITY: Okemos
STATE: Michigan
COUNTRY: USA
ZIP: 48864
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 360 KB
MEDIUM TYPE: storage
COMPUTER: Acer
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/820,980
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/424,641
FILING DATE: April 19, 1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Ian C. McLeod

REGISTRATION NUMBER: 20,931
REFERENCE/DOCKET NUMBER: Quest 4.1-156
TELECOMMUNICATION INFORMATION:
TELEPHONE: (517) 347-4100
TELEFAX: (517) 347-4103
TELEX: No. 5925388e
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 304 Amino Acids
TYPE: Amino Acid
STRANDEDNESS: Single
TOPOLOGY: Linear
MOLECULE TYPE: Peptide
US-08-820-980-4

Query Match 2.2%; Score 7; DB 2; Length 304;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 80 TVSKIAQ 86
Db 171 TVSKIAQ 177

RESULT 16
US-08-826-439-4
Sequence 4, Application US/08826439
Patent No. 5972673
GENERAL INFORMATION:
APPLICANT: Sylvain Moineau, Shirley A.
APPLICANT: Walker, Ebenezer R. Vedamuthu,
APPLICANT: and Peter A. Vandenberg
TITLE OF INVENTION: Isolated DNA Encoding
TITLE OF INVENTION: Enzyme For Phage
TITLE OF INVENTION: Resistance
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ian C. McLeod
STREET: 2190 Commons Parkway
CITY: Okemos
STATE: Michigan
COUNTRY: USA
ZIP: 48864
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 360 KB
MEDIUM TYPE: storage
COMPUTER: Acer
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/826,439
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/424,641
FILING DATE: April 19, 1995
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Ian C. McLeod
REGISTRATION NUMBER: 20,931
REFERENCE/DOCKET NUMBER: Quest 4.1-155
TELECOMMUNICATION INFORMATION:
TELEPHONE: (517) 347-4100
TELEFAX: (517) 347-4103
TELEX: No. 5972673e
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 304 Amino Acids
TYPE: Amino Acid
STRANDEDNESS: Single
TOPOLOGY: Linear
MOLECULE TYPE: Peptide
US-08-826-439-4

Query Match 2.2%; Score 7; DB 2; Length 304;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 80 TVSKIAQ 86
|||
171 TVSKIAQ 177

DB

RESULT 17
US-08-913-159-8
; Sequence 8, Application US/08913159
; Patent No. 6500109
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Plasmid-derived type II
; TITLE OF INVENTION: restriction-modification systems from *Lactococcus lactis*
; NUMBER OF SEQUENCES: 14
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/913.159
; FILING DATE:
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: DK 0179/95
; FILING DATE: 17-FEB-1995
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 304 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-913-159-8

Query Match 2.2%; Score 7; DB 4; Length 304;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 80 TVSKIAQ 86
|||
171 TVSKIAQ 177

DB

RESULT 18
US-09-147-236-11
; Sequence 11, Application US/09147236A
; Patent No. 6316251
; GENERAL INFORMATION:
; APPLICANT: TONOUCHI, Naoto
; APPLICANT: TSUCHIDA, Takayasu
; APPLICANT: YOSHINAGA, Fumihito
; APPLICANT: TAYASHI, Naoki
; APPLICANT: HAYASHI, Takahisa
; TITLE OF INVENTION: NOVEL GENE, GROUP OF GENES, AND NOVEL BETA-GLUCOSIDASE
; FILE REFERENCE: 6317-011-0PCT
; CURRENT APPLICATION NUMBER: US/09/147.236A
; EARLIER FILING DATE: 1995-04-08
; EARLIER APPLICATION NUMBER: PCT/JP97/03633
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 11
; LENGTH: 344
; TYPE: PRT
; ORGANISM: *Acetobacter xylinum*
; FEATURE:
; OTHER INFORMATION: n at positions 15741 and 15767 may be a, g, c, or
US-09-147-236-11

Query Match 2.2%; Score 7; DB 4; Length 344;
Best Local Similarity 100.0%; Pred. No. 11e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 177 QPAPVP 183
|||
172 QPAPVP 178

DB

RESULT 19
US-08-909-828-1
; Sequence 1, Application US/08909828
; Patent No. 6060646
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Tolerance of trichothecene Mycotoxins in
; TITLE OF INVENTION: Plants and Animals Through the Modification of the
; NUMBER OF SEQUENCES: 4
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/909.828
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 350 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-909-828-1

Query Match 2.2%; Score 7; DB 3; Length 350;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 159 TPVVVV 165
|||
81 TPVVVV 87

DB

RESULT 20
US-08-909-828-2
; Sequence 2, Application US/08909828
; Patent No. 6060646
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Tolerance of Trichothecene Mycotoxins in
; TITLE OF INVENTION: Plants and Animals Through the Modification of the
; NUMBER OF SEQUENCES: 4
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/909.828
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 350 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-909-828-2

Query Match 2.2%; Score 7; DB 3; Length 350;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 159 TTPVVVV 165
DB 81 TTPVVVV 87

RESULT 21
US-09-232-191-4
Sequence 4, Application US/09232191

Patent No. 6284487
GENERAL INFORMATION:
APPLICANT: Stahl, Andreas
APPLICANT: Hirsch, David J.
APPLICANT: Lodish, Harvey F.
TITLE OF INVENTION: Fatty Acid Transport Proteins
FILE REFERENCE: WHI97-21P3ME
CURRENT APPLICATION NUMBER: US/09/232,191
EARLIER APPLICATION NUMBER: 60/071,374
EARLIER FILING DATE: 1998-01-15
EARLIER APPLICATION NUMBER: 60/093,491
EARLIER FILING DATE: 1998-07-20
EARLIER APPLICATION NUMBER: 60/110,941
EARLIER FILING DATE: 1998-12-04
NUMBER OF SEQ ID NOS: 37
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 4
LENGTH: 356
TYPE: PRT
ORGANISM: Saccharomyces cerevisiae
US-09-232-191-4

Query Match 2.2%; Score 7; DB 4; Length 356;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 186 TEAPFAT 192
DB 145 TEAPFAT 151

RESULT 22
US-09-232-200-4
Sequence 4, Application US/09232200A

Patent No. 6288213
GENERAL INFORMATION:
APPLICANT: Stahl, Andreas
APPLICANT: Hirsch, David J.
APPLICANT: Lodish, Harvey F.
APPLICANT: Glimo, Ruth E.
APPLICANT: Tartaglia, Louis A.
TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
FILE REFERENCE: WHI97-21P3ME
CURRENT APPLICATION NUMBER: US/09/232,200A
EARLIER APPLICATION NUMBER: 60/071,374
EARLIER FILING DATE: 1998-01-15
EARLIER APPLICATION NUMBER: 60/093,491
EARLIER FILING DATE: 1998-07-20
EARLIER APPLICATION NUMBER: 60/110,941
EARLIER FILING DATE: 1998-12-04
NUMBER OF SEQ ID NOS: 105
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 4
LENGTH: 356
TYPE: PRT
ORGANISM: Saccharomyces cerevisiae
US-09-232-200-4

Query Match 2.2%; Score 7; DB 4; Length 356;

Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 186 TEAPFAT 192
DB 145 TEAPFAT 151

RESULT 23
US-09-232-197-4
Sequence 4, Application US/09232197A

Patent No. 6300096
GENERAL INFORMATION:
APPLICANT: Stahl, Andreas
APPLICANT: Hirsch, David J.
APPLICANT: Lodish, Harvey F.
APPLICANT: Glimo, Ruth E.
APPLICANT: Tartaglia, Louis A.
TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
FILE REFERENCE: WHI97-21P3MA
CURRENT APPLICATION NUMBER: US/09/232,197A
EARLIER APPLICATION NUMBER: 60/071,374
EARLIER FILING DATE: 1998-01-15
EARLIER APPLICATION NUMBER: 60/093,491
EARLIER FILING DATE: 1998-07-20
EARLIER APPLICATION NUMBER: 60/110,941
EARLIER FILING DATE: 1998-12-04
NUMBER OF SEQ ID NOS: 105
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 4
LENGTH: 356
TYPE: PRT
ORGANISM: Saccharomyces cerevisiae
US-09-232-197-4

Query Match 2.2%; Score 7; DB 4; Length 356;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 186 TEAPFAT 192
DB 145 TEAPFAT 151

RESULT 24
US-09-232-201-4
Sequence 4, Application US/09232201A

Patent No. 6348321
GENERAL INFORMATION:
APPLICANT: Stahl, Andreas
APPLICANT: Hirsch, David J.
APPLICANT: Lodish, Harvey F.
APPLICANT: Glimo, Ruth E.
APPLICANT: Tartaglia, Louis A.
TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
FILE REFERENCE: WHI97-21P3MC
CURRENT APPLICATION NUMBER: US/09/232,201A
EARLIER APPLICATION NUMBER: 60/071,374
EARLIER FILING DATE: 1998-01-15
EARLIER APPLICATION NUMBER: 60/093,491
EARLIER FILING DATE: 1998-07-20
EARLIER APPLICATION NUMBER: 60/110,941
EARLIER FILING DATE: 1998-12-04
NUMBER OF SEQ ID NOS: 105
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 4
LENGTH: 356
TYPE: PRT
ORGANISM: Saccharomyces cerevisiae
US-09-232-201-4

Query Match 2.2%; Score 7; DB 4; Length 356;

Query Match 2.2%; Score 7; DB 4; Length 356;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 186 TEAPFAT 192
Db 145 TEAPFAT 151

RESULT 25
US-08-149-105-15
Sequence 15, Application US/08149105
Patent No. 5538892
GENERAL INFORMATION:
APPLICANT: Donahoe, Patricia K.
APPLICANT: Gustafson, Michael
APPLICANT: He, Wei W.
APPLICANT: Wang, Xiao-Fan
TITLE OF INVENTION: TGF- TYPE I RECEPTOR
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 502 or 555X
OPERATING SYSTEM: MS-DOS (Version 5.0)
SOFTWARE: WordPerfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/149,105
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/029,673
FILING DATE: March 11, 1993
APPLICATION NUMBER: 07/853,396
FILING DATE: March 18, 1992
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00786/211001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 501
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-149-105-15

Query Match 2.2%; Score 7; DB 1; Length 501;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 STSGSGS 47
Db 187 STSGSGS 193

RESULT 26
US-08-317-847-15
Sequence 15, Application US/08317847
Patent No. 5547854
GENERAL INFORMATION:
APPLICANT: Donahoe, Patricia K.
APPLICANT: Gustafson, Michael

APPLICANT: He, Wei W.
TITLE OF INVENTION: FOUR NOVEL RECEPTORS OF THE TGF-B
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 502 or 555X
OPERATING SYSTEM: MS-DOS (Version 5.0)
SOFTWARE: WordPerfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/317,847
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/029,673
FILING DATE: March 11, 1993
APPLICATION NUMBER: 07/853,396
FILING DATE: March 18, 1992
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00786/127002
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 501
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-317-847-15

Query Match 2.2%; Score 7; DB 1; Length 501;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 STSGSGS 47
Db 187 STSGSGS 193

RESULT 27
US-09-382-256-8
Sequence 8, Application US/09382256A
Patent No. 6207814
GENERAL INFORMATION:
APPLICANT: MIYAZONO, Kohel
APPLICANT: TEN DIJKE, Peter
APPLICANT: PRANZEN, Petra
APPLICANT: YAMASHITA, Hidetoshi
APPLICANT: HELDIN, Carl-Henrik
TITLE OF INVENTION: ACTIVIN RECEPTOR LIKE KINASES, PROTEINS
HAVING SERINE THREONINE KINASE DOMAINS,
AND THEIR USE
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fulbright & Jaworski L.L.P.
STREET: 666 Fifth Avenue
CITY: New York City
STATE: New York
COUNTRY: USA
ZIP: 10103
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.25 inch, 1.44mb

```

;
;   COMPUTER: IBM PS/2
;   OPERATING SYSTEM: PC-DOS
;   SOFTWARE: Wordperfect
;
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/382,256A
;   FILING DATE: 24-Aug-1999
;   CLASSIFICATION: 514
;
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: PCT/GB93/02367
;   FILING DATE: No. 6207814ember 17, 1993
;   APPLICATION NUMBER: GB 9224057.1
;   FILING DATE: No. 6207814ember 17, 1992
;   APPLICATION NUMBER: GB 9304677.9
;   FILING DATE: March 8, 1993
;   APPLICATION NUMBER: GB 9304680.3
;   FILING DATE: March 8, 1993
;   APPLICATION NUMBER: 9311047.6
;   FILING DATE: May 28, 1993
;   APPLICATION NUMBER: 9313763.6
;   FILING DATE: July 2, 1993
;   APPLICATION NUMBER: 9316099.2
;   FILING DATE: August 3, 1993
;   APPLICATION NUMBER: 321344.5
;   FILING DATE: October 15, 1993
;
;   ATTORNEY/AGENT INFORMATION:
;   NAME: No. 6207814man D. Hanson
;   REGISTRATION NUMBER: 30,946
;   REFERENCE/DOCKET NUMBER: LUD 5298.1
;
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (212) 318-3000
;   TELEFAX: (212) 752-5958
;
;   INFORMATION FOR SEQ ID NO: 8:
;
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 505 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
;   SEQUENCE DESCRIPTION: SEQ ID NO: 8:
;
; US-09-382-256-8
;
Query Match      2.2%; Score 7; DB 4; Length 505;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy      41 STSGSGS 47
Db      187 STSGSGS 193

RESULT 28
US-09-382-256-16
; Sequence 16, Application US/09382256A
; Patent No. 6207814
; GENERAL INFORMATION:
; APPLICANT: MIYAZONO, Kohel
;             TEN DIJKE, Peter
;             FRANZEN, Petra
;             YAMASHITA, Hidetoshi
;             HELDIN, Carl-Henrik
; TITLE OF INVENTION: ACTIVIN RECEPTOR LIKE KINASES, PROTEINS
;                   HAVING SERINE THREONINE KINASE DOMAINS,
;                   AND THEIR USE
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fulbright & Jaworski L.L.P.
; STREET: 666 Fifth Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10103
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.25 inch, 1.44mb
; COMPUTER: IBM PS/2

```

```

;
;   OPERATING SYSTEM: PC-DOS
;   SOFTWARE: Wordperfect
;
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/382,256A
;   FILING DATE: 24-Aug-1999
;   CLASSIFICATION: 514
;
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: PCT/GB93/02367
;   FILING DATE: No. 6207814ember 17, 1993
;   APPLICATION NUMBER: GB 9224057.1
;   FILING DATE: No. 6207814ember 17, 1992
;   APPLICATION NUMBER: GB 9304677.9
;   FILING DATE: March 8, 1993
;   APPLICATION NUMBER: GB 9304680.3
;   FILING DATE: March 8, 1993
;   APPLICATION NUMBER: 9311047.6
;   FILING DATE: May 28, 1993
;   APPLICATION NUMBER: 9313763.6
;   FILING DATE: July 2, 1993
;   APPLICATION NUMBER: 9316099.2
;   FILING DATE: August 3, 1993
;   APPLICATION NUMBER: 321344.5
;   FILING DATE: October 15, 1993
;
;   ATTORNEY/AGENT INFORMATION:
;   NAME: No. 6207814man D. Hanson
;   REGISTRATION NUMBER: 30,946
;   REFERENCE/DOCKET NUMBER: LUD 5298.1
;
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (212) 318-3000
;   TELEFAX: (212) 752-5958
;
;   INFORMATION FOR SEQ ID NO: 16:
;
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 505 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
;   SEQUENCE DESCRIPTION: SEQ ID NO: 16:
;
; US-09-382-256-16
;
Query Match      2.2%; Score 7; DB 4; Length 505;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy      41 STSGSGS 47
Db      187 STSGSGS 193

RESULT 29
US-09-395-115-8
; Sequence 8, Application US/09395115
; Patent No. 6271365
; GENERAL INFORMATION:
; APPLICANT: Miyazono, Kohel; DiJke, Peter Ten;
;             Franzen, Petra; Yamashita, Hidetoshi; Heldin, Carl-Henrik
; TITLE OF INVENTION: Activin Receptor-Like Kinase, Proteins
;                   Having Serine Threonine Kinase Domains And Their Use
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; ZIP: 10022
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage
; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/395,115
; FILING DATE:
; CLASSIFICATION:

```

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/436,265
FILING DATE: 30-October-1995
APPLICATION NUMBER: PCT/GB93/02367
FILING DATE: 17-No. 6271365ember-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9224057.1
FILING DATE: 17-No. 6271365ember-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9304677.9
FILING DATE: 8-March-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9304680.3
FILING DATE: 8-March-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9311047.6
FILING DATE: 28-May-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9313763.6
FILING DATE: 2-July-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9316099.2
FILING DATE: 3-August-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9321344.5
FILING DATE: 15-October-1993
ATTORNEY/AGENT INFORMATION:
NAME: Kohlei, Vineet
REGISTRATION NUMBER: 37,003
REFERENCE/DOCKET NUMBER: LUD 5298
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 505 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-395-115-8

Query Match 2.2%; Score 7; DB 4; Length 505;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 STSGSGS 47
DB 187 STSGSGS 193

RESULT 30
US-09-395-115-16
Sequence 16, Application US/09395115
Patent No. 6271365
GENERAL INFORMATION:
APPLICANT: Miyazono, Kohlei, DiIke, Peter Ten;
APPLICANT: Franzén, Petter; Yamashita, Hidekoshi; Haldin, Carl-Henrik
TITLE OF INVENTION: Having Serine Threonine Kinase Domains And Their Use
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSER: Felte & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage
COMPUTER: IBM
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/395,115

FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/436,265
FILING DATE: 30-October-1995
APPLICATION NUMBER: PCT/GB93/02367
FILING DATE: 17-No. 6271365ember-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9224057.1
FILING DATE: 17-No. 6271365ember-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9304677.9
FILING DATE: 8-March-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9304680.3
FILING DATE: 8-March-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9311047.6
FILING DATE: 28-May-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9313763.6
FILING DATE: 2-July-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9316099.2
FILING DATE: 3-August-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9321344.5
FILING DATE: 15-October-1993
ATTORNEY/AGENT INFORMATION:
NAME: Kohlei, Vineet
REGISTRATION NUMBER: 37,003
REFERENCE/DOCKET NUMBER: LUD 5298
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 505 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-395-115-16

Query Match 2.2%; Score 7; DB 4; Length 505;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 STSGSGS 47
DB 187 STSGSGS 193

RESULT 31
US-08-123-934A-8
Sequence 8, Application US/08123934A
Patent No. 6291206
GENERAL INFORMATION:
APPLICANT: WOZNEY, John
APPLICANT: CELESTE, Anthony J.
APPLICANT: THIES, R. Scott
TITLE OF INVENTION: RECEPTOR PROTEINS
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSER: Genetics Institute Inc. - Legal Affairs
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: MA
COUNTRY: USA
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/123,934A
FILING DATE: 17-SEP-1993
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: LAZAR, Steven R.
REGISTRATION NUMBER: 32,618
REFERENCE/DOCKET NUMBER: 5203
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617 876 1170
TELEFAX: 617 876 5851
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 505 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-123-934A-8

Query Match 2.2%; Score 7; DB 4; Length 505;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 41 STSGSGS 47
Db 187 STSGSGS 193

RESULT 32
US-08-436-265-8
Sequence 8, Application US/08436265
Patent No. 6316217

GENERAL INFORMATION:

APPLICANT: Miyazono, Kohel; Dijke, Peter Ten;
APPLICANT: Franzen, Petra; Yamashita, Hidetoshi; Heldin, Carl-Henrik
TITLE OF INVENTION: Activin Receptor-Like Kinase, Proteins
TITLE OF INVENTION: Having Serine Threonine Kinase Domains And Their Use
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Felte & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
ZIP: 10022

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage

COMPUTER: IBM

OPERATING SYSTEM: PC-DOS

SOFTWARE: Wordperfect

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/436,265
FILING DATE: 30-October-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB93/02367
FILING DATE: 17-No. 6316217ember-1993

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9224057.1
FILING DATE: 17-No. 6316217ember-1992

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9304677.9
FILING DATE: 8-March-1993

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9304680.3
FILING DATE: 8-March-1993

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9311047.6
FILING DATE: 28-May-1993

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9313763.6
FILING DATE: 2-July-1993

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9136099.2
FILING DATE: 3-August-1993
PRIOR APPLICATION DATA: 9321344.5
APPLICATION NUMBER: 9321344.5
FILING DATE: 15-October-1993
ATTORNEY/AGENT INFORMATION:
NAME: Kohel, Vaneet
REGISTRATION NUMBER: 37,003
REFERENCE/DOCKET NUMBER: LUD 5298
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 688-3884
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 505 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-436-265-8

Query Match 2.2%; Score 7; DB 4; Length 505;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 41 STSGSGS 47
Db 187 STSGSGS 193

RESULT 33
US-08-436-265-16
Sequence 16, Application US/08436265
Patent No. 6316217

GENERAL INFORMATION:

APPLICANT: Miyazono, Kohel; Dijke, Peter Ten;
APPLICANT: Franzen, Petra; Yamashita, Hidetoshi; Heldin, Carl-Henrik
TITLE OF INVENTION: Activin Receptor-Like Kinase, Proteins
TITLE OF INVENTION: Having Serine Threonine Kinase Domains And Their Use
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Felte & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
ZIP: 10022

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage

COMPUTER: IBM

OPERATING SYSTEM: PC-DOS

SOFTWARE: Wordperfect

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/436,265
FILING DATE: 30-October-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB93/02367
FILING DATE: 17-No. 6316217ember-1993

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9224057.1
FILING DATE: 17-No. 6316217ember-1992

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9304677.9
FILING DATE: 8-March-1993

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9304680.3
FILING DATE: 8-March-1993

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9311047.6
FILING DATE: 28-May-1993

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9313763.6
FILING DATE: 2-July-1993

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9136099.2
FILING DATE: 3-August-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9321344.5
FILING DATE: 15-October-1993
ATTORNEY/AGENT INFORMATION:
NAME: Kohlei, Vineet
REGISTRATION NUMBER: 37,003
REFERENCE/DOCKET NUMBER: LUD 5298
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 505 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-436-265-16

Query Match 2.2%; Score 7; DB 4; Length 505;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 41 STSGSGS 47
|||
Db 187 STSGSGS 193

RESULT 34

US-09-679-187-8
Sequence 8, Application US/09679187
Patent No. 6331621
GENERAL INFORMATION:
APPLICANT: Miyazono, Kohlei; DiJke, Peter Ten;
APPLICANT: Franzen, Petra; Yamashita, Hidetoshi; Heldin, Carl-Henrik
TITLE OF INVENTION: Activin Receptor-Like Kinase, Proteins
TITLE OF INVENTION: Having Serine Threonine Kinase Domains And Their Use
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Felte & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage
COMPUTER: IBM
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/679,187
FILING DATE: 03-OCT-2000
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/436,265
FILING DATE: 30-October-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB93/02367
FILING DATE: 17-NO. 6331621ember-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9224057.1
FILING DATE: 17-NO. 6331621ember-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9304677.9
FILING DATE: 8-March-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9304680.3
FILING DATE: 8-March-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9311047.6
FILING DATE: 28-May-1993
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 9313763.6
FILING DATE: 2-July-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9136099.2
FILING DATE: 3-August-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9321344.5
FILING DATE: 15-October-1993
ATTORNEY/AGENT INFORMATION:
NAME: Kohlei, Vineet
REGISTRATION NUMBER: 37,003
REFERENCE/DOCKET NUMBER: LUD 5298
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 505 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-679-187-8

Query Match 2.2%; Score 7; DB 4; Length 505;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 41 STSGSGS 47
|||
Db 187 STSGSGS 193

RESULT 35

US-09-679-187-16
Sequence 16, Application US/09679187
Patent No. 6331621
GENERAL INFORMATION:
APPLICANT: Miyazono, Kohlei; DiJke, Peter Ten;
APPLICANT: Franzen, Petra; Yamashita, Hidetoshi; Heldin, Carl-Henrik
TITLE OF INVENTION: Activin Receptor-Like Kinase, Proteins
TITLE OF INVENTION: Having Serine Threonine Kinase Domains And Their Use
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Felte & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage
COMPUTER: IBM
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/679,187
FILING DATE: 03-OCT-2000
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/436,265
FILING DATE: 30-October-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB93/02367
FILING DATE: 17-NO. 6331621ember-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9224057.1
FILING DATE: 17-NO. 6331621ember-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9304677.9
FILING DATE: 8-March-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9304680.3
FILING DATE: 8-March-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9311047.6
FILING DATE: 28-May-1993
PRIOR APPLICATION DATA:

;; FILING DATE: 28-May-1993
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 9313763.6
;; FILING DATE: 2-July-1993
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 9136099.2
;; FILING DATE: 3-August-1993
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 9321344.5
;; FILING DATE: 15-October-1993
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Kohlei, Vineet
;; REGISTRATION NUMBER: 37,003
;; REFERENCE/DOCKET NUMBER: LUD 5298
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (212) 688-9200
;; TELEFAX: (212) 838-3884
;; INFORMATION FOR SEQ ID NO: 16:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 505 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-09-679-187-16.
Query Match 2.2%; Score 7; DB 4; Length 505;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 41 STSGSGS 47
|||
Db 187 STSGSGS 193
RESULT 36
PCT-US94-10080-8
; Sequence 8, Application PC/TUS9410080
; GENERAL INFORMATION:
; APPLICANT: GENETICS INSTITUTE, INC.
; TITLE OF INVENTION: RECEPTOR PROTEINS
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Genetics Institute Inc. - Legal Affairs
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/10080
; FILING DATE: HERewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/123,934
; FILING DATE: 17-SEP-1993
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: LAZAR, Steven R
; REGISTRATION NUMBER: 32,618
; REFERENCE/DOCKET NUMBER: 5203-PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8260
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 505 amino acids
; TYPE: amino acid
; TOPOLOGY: linear

;; MOLECULE TYPE: protein
PCT-US94-10080-8
Query Match 2.2%; Score 7; DB 5; Length 505;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 41 STSGSGS 47
|||
Db 187 STSGSGS 193
RESULT 37
US-08-469-412A-7
; Sequence 7, Application US/08469412A
; Patent No. 5856125
; GENERAL INFORMATION:
; APPLICANT: Mavrothalasitis, George J.
; APPLICANT: Blair, Donald G.
; APPLICANT: Fisher, Robert J.
; APPLICANT: Beal Jr., Gregory J.
; APPLICANT: Athanasios, Metropi A.
; APPLICANT: Sgouras, Dionysios N.
; TITLE OF INVENTION: The ERF Genetic Locus and Its Products
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,412A
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Garrett-Mackowski, Eugenia
; REGISTRATION NUMBER: 37,330
; REFERENCE/DOCKET NUMBER: 015280-229000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 543 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..543
; OTHER INFORMATION: /note="murine ERF amino acid sequence
; OTHER INFORMATION: (first 8 amino acids from first exon not
; ; OTHER INFORMATION: included)"
US-08-469-412A-7
Query Match 2.2%; Score 7; DB 2; Length 543;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 51 SSGSGLA 57
|||
Db 398 SSGSGLA 404
RESULT 38

US-09-021-715-7
; Sequence 7, Application US/09021715
; Patent No. 6194547
; GENERAL INFORMATION:
; APPLICANT: Macrothallacis, George J.
; Blair, Donald G.
; Fisher, Robert J.
; Beal Jr., Gregory J.
; Athanasiou, Meropi A.
; Sgouras, Dionysios N.
; TITLE OF INVENTION: The ERF Genetic Locus and Its Products
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/021,715
; FILING DATE: 10-Feb-1998
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Garrett-Wackowski, Eugenia
; REGISTRATION NUMBER: 37,330
; REFERENCE/DOCKET NUMBER: 015280-229000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 543 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; FEATURE:
; MOLECULE TYPE: protein
; NAME/KEY: Protein
; LOCATION: 1-543
; OTHER INFORMATION: /note="murine ERF amino acid sequence
; (first 8 amino acids from first exon not
; included)
; SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-09-021-715-7
Query Match 2.2%; Score 7; DB 4; Length 543;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 51 SSGGGGLA 57
DB 398 SSGGGGLA 404
RESULT 39
US-09-232-200-89
; Sequence 89, Application US/09232200A
; Patent No. 6288213
; GENERAL INFORMATION:
; APPLICANT: Stahl, Andreas
; APPLICANT: Hirsch, David J.
; APPLICANT: Lodish, Harvey F.
; APPLICANT: Gimeno, Ruth E.
; APPLICANT: Tartaglia, Louis A.
; TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
; FILE REFERENCE: WHI97-21P3MB
; CURRENT APPLICATION NUMBER: US/09/232,200A

; CURRENT FILING DATE: 1998-01-14
; EARLIER APPLICATION NUMBER: 60/071,374
; EARLIER FILING DATE: 1998-01-15
; EARLIER APPLICATION NUMBER: 60/093,491
; EARLIER FILING DATE: 1998-07-20
; EARLIER APPLICATION NUMBER: 60/110,941
; EARLIER FILING DATE: 1998-12-04
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 89
; LENGTH: 623
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-232-200-89

Query Match 2.2%; Score 7; DB 4; Length 623;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 186 TEAPFAT 192
DB 398 TEAPFAT 404

RESULT 40
US-09-232-200-98
; Sequence 98, Application US/09232200A
; Patent No. 6288213
; GENERAL INFORMATION:
; APPLICANT: Stahl, Andreas
; APPLICANT: Hirsch, David J.
; APPLICANT: Lodish, Harvey F.
; APPLICANT: Gimeno, Ruth E.
; APPLICANT: Tartaglia, Louis A.
; TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
; FILE REFERENCE: WHI97-21P3MB
; CURRENT APPLICATION NUMBER: US/09/232,200A
; CURRENT FILING DATE: 1999-01-14
; EARLIER APPLICATION NUMBER: 60/071,374
; EARLIER FILING DATE: 1998-01-15
; EARLIER APPLICATION NUMBER: 60/093,491
; EARLIER FILING DATE: 1998-07-20
; EARLIER APPLICATION NUMBER: 60/110,941
; EARLIER FILING DATE: 1998-12-04
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 98
; LENGTH: 623
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-232-200-98

Query Match 2.2%; Score 7; DB 4; Length 623;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 186 TEAPFAT 192
DB 398 TEAPFAT 404

Search completed: July 8, 2003, 11:22:01
Job time: 42 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 8, 2003, 11:21:03 ; Search time 54 Seconds
(without alignments)
694.207 Million cell updates/sec

Title: US-10-018-706-2
Perfect score: 322
Sequence: 1 MTVTAINSONQKPIKRLGL.....LFEFRISRMGYVPLTLX 322

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 445758 seqs, 11641973 residues

Word size : 0

Total number of hits satisfying chosen parameters: 445758

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database :

Published Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PC7_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/PC7US_PUBCOMB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US05_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8	2.5	223	9	US-09-738-626-3830
2	7	2.2	13	10	US-09-944-411-9
3	7	2.2	13	10	US-09-759-352-44
4	7	2.2	208	9	US-09-533-029-70
5	7	2.2	208	9	US-10-286-264-72
6	7	2.2	208	9	US-10-180-375-195
7	7	2.2	213	10	US-09-771-161A-109
8	7	2.2	251	9	US-10-125-540-300
9	7	2.2	251	10	US-09-764-870-300
10	7	2.2	343	9	US-10-156-761-14239
11	7	2.2	356	10	US-09-943-671-4
12	7	2.2	375	9	US-10-156-761-8098
13	7	2.2	490	9	US-10-156-761-13366
14	7	2.2	496	9	US-10-081-872-212
15	7	2.2	505	10	US-09-771-161A-200
16	7	2.2	505	10	US-09-903-068-8
17	7	2.2	505	10	US-09-903-068-16
18	7	2.2	505	10	US-09-874-628-8
19	7	2.2	806	9	US-09-738-626-5890

20	7	2.2	843	9	US-09-978-303-25	Sequence 25, Appl
21	7	2.2	1336	9	US-10-278-173-128	Sequence 128, Appl
22	7	2.2	1336	9	US-09-934-455-22	Sequence 22, Appl
23	7	2.2	2344	10	US-09-815-242-12713	Sequence 12713, A
24	7	2.2	2796	9	US-09-870-759-114	Sequence 114, Appl
25	7	2.2	3069	9	US-09-712-363-246	Sequence 246, Appl
26	6	1.9	14	9	US-09-956-086-7	Sequence 7, Appl1
27	6	1.9	14	9	US-09-999-745-58	Sequence 58, Appl
28	6	1.9	14	9	US-09-999-745-60	Sequence 60, Appl
29	6	1.9	14	9	US-09-956-087-7	Sequence 7, Appl1
30	6	1.9	14	9	US-09-554-000-42	Sequence 42, Appl
31	6	1.9	14	9	US-09-554-000-44	Sequence 44, Appl
32	6	1.9	14	9	US-09-792-793A-4	Sequence 4, Appl1
33	6	1.9	14	9	US-09-792-793A-6	Sequence 6, Appl1
34	6	1.9	14	9	US-10-158-238-9	Sequence 9, Appl1
35	6	1.9	14	9	US-10-247-279-3	Sequence 3, Appl1
36	6	1.9	14	9	US-10-052-942-6	Sequence 6, Appl1
37	6	1.9	14	10	US-09-883-777-13	Sequence 13, Appl
38	6	1.9	14	10	US-09-987-456-11	Sequence 11, Appl
39	6	1.9	16	10	US-09-883-777-14	Sequence 14, Appl
40	6	1.9	16	10	US-09-883-777-15	Sequence 15, Appl
41	6	1.9	17	9	US-10-091-724-42	Sequence 42, Appl
42	6	1.9	18	9	US-09-956-086-8	Sequence 8, Appl1
43	6	1.9	18	9	US-09-956-086-10	Sequence 10, Appl
44	6	1.9	18	9	US-09-999-745-59	Sequence 59, Appl
45	6	1.9	18	9	US-09-999-745-61	Sequence 61, Appl
46	6	1.9	18	9	US-09-956-087-8	Sequence 8, Appl1
47	6	1.9	18	9	US-09-956-087-10	Sequence 10, Appl
48	6	1.9	18	9	US-09-554-000-43	Sequence 43, Appl
49	6	1.9	18	9	US-09-554-000-45	Sequence 45, Appl
50	6	1.9	18	9	US-09-792-793A-5	Sequence 5, Appl1
51	6	1.9	18	9	US-09-792-793A-7	Sequence 7, Appl1
52	6	1.9	18	9	US-09-866-538-26	Sequence 26, Appl
53	6	1.9	18	9	US-10-158-238-10	Sequence 10, Appl
54	6	1.9	18	9	US-10-158-238-11	Sequence 11, Appl
55	6	1.9	18	9	US-10-121-258-18	Sequence 38, Appl
56	6	1.9	18	9	US-10-247-279-5	Sequence 4, Appl1
57	6	1.9	18	9	US-10-247-279-5	Sequence 5, Appl1
58	6	1.9	19	10	US-09-168-474-19	Sequence 19, Appl
59	6	1.9	22	9	US-09-899-235-20	Sequence 20, Appl
60	6	1.9	22	9	US-10-121-258-19	Sequence 19, Appl
61	6	1.9	23	9	US-10-023-282-800	Sequence 800, Appl
62	6	1.9	28	9	US-10-210-895-9	Sequence 9, Appl1
63	6	1.9	32	10	US-09-864-761-38752	Sequence 38752, A
64	6	1.9	33	10	US-09-864-761-38943	Sequence 38943, A
65	6	1.9	34	9	US-10-106-638-8038	Sequence 8038, Appl
66	6	1.9	41	9	US-10-001-876-124	Sequence 124, Appl
67	6	1.9	42	9	US-09-925-299-1393	Sequence 1393, Appl
68	6	1.9	42	10	US-09-925-299-1393	Sequence 1393, Appl
69	6	1.9	44	9	US-10-023-282-799	Sequence 799, Appl
70	6	1.9	48	10	US-09-864-761-45369	Sequence 45369, A
71	6	1.9	54	9	US-09-983-802-424	Sequence 224, Appl
72	6	1.9	55	10	US-09-864-761-40446	Sequence 40446, A
73	6	1.9	62	9	US-10-106-638-5547	Sequence 6547, Appl
74	6	1.9	63	9	US-09-738-626-6587	Sequence 6587, Appl
75	6	1.9	65	9	US-10-023-282-449	Sequence 949, Appl
76	6	1.9	69	10	US-09-864-761-45687	Sequence 45687, A
77	6	1.9	76	10	US-09-168-474-6	Sequence 45905, A
78	6	1.9	83	9	US-10-092-154-599	Sequence 599, Appl
79	6	1.9	83	9	US-09-764-847-559	Sequence 559, Appl
80	6	1.9	89	9	US-10-156-761-12606	Sequence 12606, A
81	6	1.9	101	9	US-09-738-626-6090	Sequence 6090, Appl
82	6	1.9	104	10	US-09-867-550-1450	Sequence 1450, Appl
83	6	1.9	104	10	US-09-755-100-7	Sequence 7, Appl1
84	6	1.9	104	10	US-10-023-282-745	Sequence 745, Appl
85	6	1.9	105	10	US-09-864-761-42752	Sequence 42752, A
86	6	1.9	105	10	US-09-925-301-986	Sequence 986, Appl
87	6	1.9	106	10	US-09-734-569-136	Sequence 136, Appl
88	6	1.9	107	10	US-09-740-002-20	Sequence 20, Appl
89	6	1.9	116	9	US-10-073-961-254	Sequence 254, Appl
90	6	1.9	116	10	US-09-764-887-254	Sequence 254, Appl
91	6	1.9	119	10	US-09-820-893-50	Sequence 50, Appl

677	1.9	1426	9	US-10-323-579-15	Sequence 15, Appl	750	1.6	10	9	US-09-572-404B-2432	Sequence 2432, Ap
678	1.9	1541	9	US-09-838-896-3	Sequence 3, Appl	751	1.6	10	9	US-09-572-404B-2436	Sequence 2436, Ap
679	1.9	1541	9	US-10-088-805-3	Sequence 3, Appl	752	1.6	10	9	US-09-572-404B-2461	Sequence 2461, Ap
680	1.9	1626	9	US-10-183-990-11	Sequence 11, Appl	753	1.6	10	9	US-09-572-404B-2552	Sequence 2552, Ap
681	1.9	1681	10	US-09-820-653-3	Sequence 8, Appl	754	1.6	10	9	US-09-572-404B-2559	Sequence 2559, Ap
682	1.9	1734	10	US-09-862-927-81	Sequence 81, Appl	755	1.6	10	9	US-09-572-404B-2561	Sequence 2561, Ap
683	1.9	1776	10	US-09-841-132-179	Sequence 19, Appl	756	1.6	10	9	US-09-572-404B-2620	Sequence 2620, Ap
684	1.9	1848	9	US-09-830-996-6	Sequence 6, Appl	757	1.6	10	9	US-09-572-404B-2622	Sequence 2622, Ap
685	1.9	1848	9	US-10-080-505-6	Sequence 6, Appl	758	1.6	10	9	US-09-572-404B-2660	Sequence 2660, Ap
686	1.9	2061	9	US-10-060-036-181	Sequence 181, Appl	759	1.6	10	9	US-09-572-404B-2694	Sequence 2694, Ap
687	1.9	2364	9	US-10-156-761-7834	Sequence 7834, Ap	760	1.6	10	9	US-09-572-404B-2894	Sequence 2894, Ap
688	1.9	2402	9	US-10-203-224-2	Sequence 2, Appl	761	1.6	10	9	US-09-572-404B-3224	Sequence 3224, Ap
689	1.9	2476	9	US-09-824-574-7	Sequence 7, Appl	762	1.6	10	9	US-09-572-404B-3270	Sequence 3270, Ap
690	1.9	2502	10	US-09-772-316-1	Sequence 1, Appl	763	1.6	10	9	US-09-572-404B-3272	Sequence 3272, Ap
691	1.9	2655	10	US-09-864-761-34248	Sequence 34248, A	764	1.6	10	9	US-09-572-404B-3283	Sequence 3283, Ap
692	1.9	2749	9	US-10-227-338-4	Sequence 4, Appl	765	1.6	10	9	US-09-572-404B-3282	Sequence 3282, Ap
693	1.9	3353	10	US-09-888-615-64	Sequence 64, Appl	766	1.6	10	9	US-09-572-404B-3284	Sequence 3284, Ap
694	1.9	3571	10	US-10-150-821-2	Sequence 2, Appl	767	1.6	10	9	US-09-572-404B-3286	Sequence 3286, Ap
695	1.9	3571	10	US-09-911-842-2	Sequence 2, Appl	768	1.6	10	9	US-09-572-404B-3370	Sequence 3370, Ap
696	1.9	3594	9	US-10-150-821-4	Sequence 4, Appl	769	1.6	10	9	US-09-572-404B-3450	Sequence 3450, Ap
697	1.9	3594	9	US-09-911-842-4	Sequence 4, Appl	770	1.6	10	9	US-09-572-404B-3452	Sequence 3452, Ap
698	1.9	3664	9	US-10-177-293-423	Sequence 423, Appl	771	1.6	10	9	US-09-572-404B-3528	Sequence 3528, Ap
699	1.9	4840	9	US-10-156-761-10435	Sequence 10435, A	772	1.6	10	9	US-09-572-404B-3532	Sequence 3532, Ap
700	1.6	5	9	US-10-177-293-423	Sequence 423, Appl	773	1.6	10	9	US-09-572-404B-3536	Sequence 3536, Ap
701	1.6	7	9	US-10-093-200A-12	Sequence 6, Appl	774	1.6	10	9	US-09-572-404B-3566	Sequence 3566, Ap
702	1.6	7	9	US-10-093-200A-12	Sequence 12, Appl	775	1.6	10	9	US-09-572-404B-3566	Sequence 3566, Ap
703	1.6	8	9	US-09-765-555-61	Sequence 61, Appl	776	1.6	10	9	US-09-572-404B-3655	Sequence 3655, Ap
704	1.6	8	9	US-09-813-153-237	Sequence 237, Appl	777	1.6	10	9	US-09-572-404B-3657	Sequence 3657, Ap
705	1.6	8	9	US-10-104-919-48	Sequence 48, Appl	778	1.6	10	9	US-09-572-404B-3676	Sequence 3676, Ap
706	1.6	8	9	US-09-780-658A-24	Sequence 24, Appl	779	1.6	10	9	US-09-572-404B-3677	Sequence 3677, Ap
707	1.6	10	9	US-10-040-572A-152	Sequence 152, Appl	780	1.6	10	9	US-09-572-404B-3678	Sequence 3678, Ap
708	1.6	10	9	US-09-572-404B-352	Sequence 152, Appl	781	1.6	10	9	US-09-572-404B-3679	Sequence 3679, Ap
709	1.6	10	9	US-09-572-404B-356	Sequence 256, Appl	782	1.6	10	9	US-09-572-404B-3680	Sequence 3680, Ap
710	1.6	10	9	US-09-572-404B-453	Sequence 453, Appl	783	1.6	10	9	US-09-572-404B-3681	Sequence 3681, Ap
711	1.6	10	9	US-09-572-404B-453	Sequence 453, Appl	784	1.6	10	9	US-09-572-404B-3682	Sequence 3682, Ap
712	1.6	10	9	US-09-572-404B-457	Sequence 457, Appl	785	1.6	10	9	US-09-572-404B-3683	Sequence 3683, Ap
713	1.6	10	9	US-09-572-404B-459	Sequence 459, Appl	786	1.6	10	9	US-09-572-404B-3700	Sequence 3700, Ap
714	1.6	10	9	US-09-572-404B-471	Sequence 471, Appl	787	1.6	10	9	US-09-572-404B-4186	Sequence 4186, Ap
715	1.6	10	9	US-09-572-404B-473	Sequence 473, Appl	788	1.6	10	9	US-09-572-404B-4187	Sequence 4187, Ap
716	1.6	10	9	US-09-572-404B-475	Sequence 475, Appl	789	1.6	10	9	US-09-572-404B-4189	Sequence 4189, Ap
717	1.6	10	9	US-09-572-404B-493	Sequence 493, Appl	790	1.6	10	9	US-09-572-404B-4189	Sequence 4189, Ap
718	1.6	10	9	US-09-572-404B-497	Sequence 497, Appl	791	1.6	10	9	US-09-572-404B-4189	Sequence 4189, Ap
719	1.6	10	9	US-09-572-404B-501	Sequence 501, Appl	792	1.6	10	9	US-09-572-404B-4189	Sequence 4189, Ap
720	1.6	10	9	US-09-572-404B-503	Sequence 503, Appl	793	1.6	10	9	US-09-572-404B-4189	Sequence 4189, Ap
721	1.6	10	9	US-09-572-404B-503	Sequence 503, Appl	794	1.6	10	9	US-09-572-404B-4189	Sequence 4189, Ap
722	1.6	10	9	US-09-572-404B-505	Sequence 505, Appl	795	1.6	10	9	US-09-572-404B-4189	Sequence 4189, Ap
723	1.6	10	9	US-09-572-404B-507	Sequence 507, Appl	796	1.6	10	9	US-09-572-404B-4189	Sequence 4189, Ap
724	1.6	10	9	US-09-572-404B-515	Sequence 515, Appl	797	1.6	10	9	US-09-572-404B-4189	Sequence 4189, Ap
725	1.6	10	9	US-09-572-404B-515	Sequence 515, Appl	798	1.6	10	9	US-09-572-404B-4189	Sequence 4189, Ap
726	1.6	10	9	US-09-572-404B-517	Sequence 517, Appl	799	1.6	10	9	US-09-572-404B-4189	Sequence 4189, Ap
727	1.6	10	9	US-09-572-404B-517	Sequence 517, Appl	800	1.6	10	9	US-09-572-404B-4189	Sequence 4189, Ap
728	1.6	10	9	US-09-572-404B-543	Sequence 543, Appl	801	1.6	10	9	US-09-572-404B-4189	Sequence 4189, Ap
729	1.6	10	9	US-09-572-404B-543	Sequence 543, Appl	802	1.6	10	9	US-09-572-404B-4189	Sequence 4189, Ap
730	1.6	10	9	US-09-572-404B-545	Sequence 545, Appl	803	1.6	10	9	US-09-572-404B-4189	Sequence 4189, Ap
731	1.6	10	9	US-09-572-404B-1051	Sequence 1051, Appl	804	1.6	10	9	US-09-572-404B-4189	Sequence 4189, Ap
732	1.6	10	9	US-09-572-404B-1051	Sequence 1051, Appl	805	1.6	10	9	US-09-572-404B-4189	Sequence 4189, Ap
733	1.6	10	9	US-09-572-404B-1051	Sequence 1051, Appl	806	1.6	10	9	US-09-572-404B-4189	Sequence 4189, Ap
734	1.6	10	9	US-09-572-404B-1081	Sequence 1081, Appl	807	1.6	10	9	US-09-572-404B-4189	Sequence 4189, Ap
735	1.6	10	9	US-09-572-404B-1083	Sequence 1083, Appl	808	1.6	10	9	US-09-572-404B-4189	Sequence 4189, Ap
736	1.6	10	9	US-09-572-404B-1083	Sequence 1083, Appl	809	1.6	10	9	US-09-572-404B-4189	Sequence 4189, Ap
737	1.6	10	9	US-09-572-404B-1087	Sequence 1087, Appl	810	1.6	10	9	US-09-572-404B-4189	Sequence 4189, Ap
738	1.6	10	9	US-09-572-404B-1095	Sequence 1095, Appl	811	1.6	10	9	US-09-572-404B-4189	Sequence 4189, Ap
739	1.6	10	9	US-09-572-404B-1103	Sequence 1103, Appl	812	1.6	10	9	US-09-572-404B-4189	Sequence 4189, Ap
740	1.6	10	9	US-09-572-404B-1105	Sequence 1105, Appl	813	1.6	10	9	US-09-572-404B-4189	Sequence 4189, Ap
741	1.6	10	9	US-09-572-404B-1206	Sequence 1206, Appl	814	1.6	10	9	US-09-572-404B-4189	Sequence 4189, Ap
742	1.6	10	9	US-09-572-404B-1208	Sequence 1208, Appl	815	1.6	10	9	US-09-572-404B-4189	Sequence 4189, Ap
743	1.6	10	9	US-09-572-404B-1210	Sequence 1210, Appl	816	1.6	10	9	US-09-572-404B-4189	Sequence 4189, Ap
744	1.6	10	9	US-09-572-404B-1376	Sequence 1376, Appl	817	1.6	10	9	US-09-572-404B-4189	Sequence 4189, Ap
745	1.6	10	9	US-09-572-404B-1378	Sequence 1378, Appl	818	1.6	10	9	US-09-572-404B-4189	Sequence 4189, Ap
746	1.6	10	9	US-09-572-404B-1380	Sequence 1380, Appl	819	1.6	10	9	US-09-572-404B-4189	Sequence 4189, Ap
747	1.6	10	9	US-09-572-404B-1440	Sequence 1440, Appl	820	1.6	10	9	US-09-572-404B-4189	Sequence 4189, Ap
748	1.6	10	9	US-09-572-404B-1510	Sequence 1510, Appl	821	1.6	10	9	US-09-572-404B-4189	Sequence 4189, Ap
749	1.6	10	9	US-09-572-404B-1574	Sequence 1574, Appl	822	1.6	10	9	US-09-572-404B-4189	Sequence 4189, Ap

969 5 1.6 26 9 US-09-969-730-272 Sequence 272, App
970 5 1.6 26 9 US-10-023-282-1220 Sequence 1220, App
971 5 1.6 26 9 US-10-213-512-177 Sequence 177, Appl
972 5 1.6 26 10 US-09-864-761-40782 Sequence 40782, A
973 5 1.6 26 10 US-09-864-761-41159 Sequence 41159, A
974 5 1.6 26 10 US-09-864-761-43426 Sequence 43426, A
975 5 1.6 26 10 US-09-921-397-7 Sequence 7, Appl1
976 5 1.6 26 10 US-09-921-397-35 Sequence 35, Appl1
977 5 1.6 26 10 US-09-071-838-17 Sequence 17, Appl
978 5 1.6 27 9 US-09-745-078A-11 Sequence 11, Appl
979 5 1.6 27 9 US-10-062-548-165 Sequence 165, App
980 5 1.6 27 10 US-09-864-761-40237 Sequence 40237, A
981 5 1.6 27 10 US-09-864-761-46273 Sequence 46273, A
982 5 1.6 27 10 US-09-939-980-531 Sequence 531, App
983 5 1.6 27 10 US-09-825-012-30 Sequence 30, Appl
984 5 1.6 28 9 US-09-999-724-28 Sequence 28, Appl
985 5 1.6 28 9 US-09-745-078A-10 Sequence 10, Appl
986 5 1.6 28 10 US-09-921-397-19 Sequence 19, Appl
987 5 1.6 29 9 US-09-745-078A-9 Sequence 9, Appl1
988 5 1.6 29 9 US-10-091-504-674 Sequence 674, App
989 5 1.6 29 10 US-09-864-761-48468 Sequence 48468, A
990 5 1.6 29 10 US-09-764-869-674 Sequence 674, App
991 5 1.6 29 10 US-09-935-291A-27 Sequence 27, Appl
992 5 1.6 29 10 US-09-935-291A-28 Sequence 28, Appl
993 5 1.6 29 10 US-09-935-291A-29 Sequence 29, Appl
994 5 1.6 29 10 US-09-935-291A-30 Sequence 30, Appl
995 5 1.6 29 10 US-08-335-281A-31 Sequence 31, Appl
996 5 1.6 29 10 US-08-335-281A-32 Sequence 32, Appl
997 5 1.6 29 10 US-08-335-281A-33 Sequence 33, Appl
998 5 1.6 29 10 US-08-935-281A-34 Sequence 34, Appl
999 5 1.6 29 10 US-09-902-741-5 Sequence 5, Appl1
1000 5 1.6 30 9 US-09-745-078A-8 Sequence 8, Appl1

ALIGNMENTS

RESULT 1
US-09-738-626-3830
; Sequence 3830, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 2A3-125
; CURRENT APPLICATION NUMBER: US/09/738, 626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO: 3830
; LENGTH: 223
; TYPE: PR1
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-3830
Query Match 2.5%; Score 8; DB 9; Length 223;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 135 HTSPVAV 142
Db 130 HTSPVAV 137

RESULT 2
US-09-944-411-9
; Sequence 9, Application US/09944411
; Patent No. US20020106792A1
; GENERAL INFORMATION:
; APPLICANT: FINER, MITCHELL H.
; DULL, THOMAS J.
; ZSEBO, KRISTINA M.
; COOKE, KEEGAN
; FARRSON, DEBORAH A.
; TITLE OF INVENTION: METHOD FOR PRODUCTION OF HIGH TITER
; VIRUS AND HIGH EFFICIENCY RETROVIRAL MEDIATED TRANSDUCTION
; OF MAMMALIAN CELLS
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESSES:
; ADDRESSSEE: CELL GENESYS, INC.
; STREET: 322 LAKESIDE DRIVE
; CITY: FOSTER CITY
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/944, 411
; FILING DATE: 04-Sep-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/914, 893
; FILING DATE: <Unknown>
; APPLICATION NUMBER: US 08/258, 152
; FILING DATE: 10-JUN-1998
; APPLICATION NUMBER: US 08/076, 299
; FILING DATE: 11-JUN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: KRUPEN, KAREN I.
; REGISTRATION NUMBER: 34,647
; REFERENCE/DOCKET NUMBER: CELL 13.3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-358-9600 X131
; TELEFAX: 415-349-7392
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 13 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 9:
US-09-944-411-9
Query Match 2.2%; Score 7; DB 10; Length 13;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 41 STSGSGS 47
Db 2 STSGSGS 8
RESULT 3
US-09-759-352-44
; Sequence 44, Application US/09759352
; Patent No. US2002011474A1

GENERAL INFORMATION:
APPLICANT: CAPON, DANIEL J
WEISS, ARTHUR
IRVING, BRYAN A
ROBERTS, MARGO R
ZSEBO, KRISZTINA
TITLE OF INVENTION: CHIMERIC CHAINS FOR RECEPTOR-ASSOCIATED
SIGNAL TRANSDUCTION PATHWAYS
NUMBER OF SEQUENCES: 51
CORRESPONDENCE ADDRESS:
ADDRESSEE: CELL GENESYS, INC.
STREET: 322 LAKESIDE DRIVE
CITY: FOSTER CITY
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94404
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/759,352
FILING DATE: 16-Jan-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/567,393
FILING DATE: 01-DEC-1995
APPLICATION NUMBER: US 08/475,442
FILING DATE: 07-JUN-1995
APPLICATION NUMBER: US 08/238,405
FILING DATE: 05-MAY-1994
APPLICATION NUMBER: US 07/988,194
FILING DATE: 09-DEC-1992
APPLICATION NUMBER: US 07/627,643
FILING DATE: 13-DEC-1990
APPLICATION NUMBER: WO PCT/US91/09431
FILING DATE: 12-DEC-1991
ATTORNEY/AGENT INFORMATION:
NAME: KRUPEN, KAREN I
REGISTRATION NUMBER: 34,647
REFERENCE/DOCKET NUMBER: CELL 5.13
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415)358-9600 X131
TELEFAX: (415)349-7392
INFORMATION FOR SEQ ID NO: 44:
SEQUENCE CHARACTERISTICS:
LENGTH: 13 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 44:
US-09-759-352-44
Query Match 2.2%; Score 7; DB 10; Length 13;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 41 STSGSGS 47
Db 2 STSGSGS 8
RESULT 4
US-09-533-029-70
Sequence 70, Application US/09533029
Publication No. US20030046723A1
GENERAL INFORMATION:
APPLICANT: Heard, Jacqueline
Broun, Pierre
APPLICANT: Broun, Pierre
APPLICANT: Riechmann, Jose-Luis
APPLICANT: Keddle, James

APPLICANT: Pineda, Omaira
APPLICANT: Adam, Luc
APPLICANT: Samaha, Raymond
APPLICANT: Zhang, James
APPLICANT: Yu, Guo-Liang
APPLICANT: Ratcliffe, Oliver
APPLICANT: Pilgrim, Marsha
APPLICANT: Jiang, Cai-Zhong
APPLICANT: Reuber, Lynne
TITLE OF INVENTION: DISEASE-INDUCED POLYNUCLEOTIDES
FILE REFERENCE: MBI-010
CURRENT APPLICATION NUMBER: US/09/533,029
CURRENT FILING DATE: 2000-03-22
EARLIER APPLICATION NUMBER: 60/125,814
EARLIER FILING DATE: 1999-03-23
NUMBER OF SEQ ID NOS: 121
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 70
LENGTH: 208
TYPE: PRT
ORGANISM: Arabidopsis thaliana
FEATURE:
OTHER INFORMATION: G620
US-09-533-029-70
Query Match 2.2%; Score 7; DB 9; Length 208;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 314 YVDPLTV 320
Db 105 YVDPLTV 111
RESULT 5
US-10-286-264-72
Sequence 72, Application US/10286264
Publication No. US20030093837A1
GENERAL INFORMATION:
APPLICANT: Keddle, James
APPLICANT: Riechmann, Jose-Luis
APPLICANT: Ratcliffe, Oliver
APPLICANT: Zhang, James
APPLICANT: Jiang, Cai-Zhong
APPLICANT: Pineda, Omaira
APPLICANT: Heard, Jacqueline
APPLICANT: Yu, Guo-Liang
APPLICANT: Adam, Luc
APPLICANT: Broun, Pierre
APPLICANT: Reuber, Lynne
APPLICANT: Samaha, Raymond
APPLICANT: Pineda, Omaira
TITLE OF INVENTION: POLYNUCLEOTIDES FOR SEED TRAIT ALTERATION
FILE REFERENCE: MBI-008
CURRENT APPLICATION NUMBER: US/10/286,264
CURRENT FILING DATE: 2002-11-01
PRIOR APPLICATION NUMBER: 60/125,814
PRIOR FILING DATE: 1999-03-23
NUMBER OF SEQ ID NOS: 165
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 72
LENGTH: 208
TYPE: PRT
ORGANISM: Arabidopsis thaliana
FEATURE:
OTHER INFORMATION: G620
US-10-286-264-72
Query Match 2.2%; Score 7; DB 9; Length 208;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 314 YVDPLTV 320

Db 105 YVDPLTV 111

RESULT 6
US-10-180-375-195
Sequence 195, Application US/10180375
Publication No. US20030126638A1
GENERAL INFORMATION:
APPLICANT: Allen, William B.
APPLICANT: Cahoon, Rebecca E.
APPLICANT: Famodu, Omolayo O.
APPLICANT: Harveill, Leslie T.
APPLICANT: Helentjaris, Timothy
APPLICANT: Li, Changjiang
APPLICANT: Lowe, Keith
APPLICANT: Oliveira, Igor Cunha
APPLICANT: Shen, Bo
APPLICANT: Tarczynski, Mitchell C.
TITLE OF INVENTION: Alteration Of Oil Traits In Plants
FILE REFERENCE: B81458 US NAI
CURRENT APPLICATION NUMBER: US/10/180,375
CURRENT FILING DATE: 2002-06-26
NUMBER OF SEQ ID NOS: 222
SOFTWARE: Microsoft Office 97
SEQ ID NO 195
LENGTH: 208
TYPE: PRT
ORGANISM: Arabidopsis thaliana
FEATURE:
NAME/KEY: misc_feature
LOCATION:
OTHER INFORMATION: gi 6552738
US-10-180-375-195

Query Match 2.2%; Score 7; DB 9; Length 208;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 314 YVDPLTV 320
Db 105 YVDPLTV 111

RESULT 7
US-09-771-161A-109
Sequence 109, Application US/09771161A
Patent No. US20020110811A1
GENERAL INFORMATION:
APPLICANT: LEVINE, et al.
TITLE OF INVENTION: VARIANTS OF PROTEIN KINASES
FILE REFERENCE: 802620-2005.1
CURRENT APPLICATION NUMBER: US/09/771,161A
CURRENT FILING DATE: 2001-01-26
PRIOR APPLICATION NUMBER: 09/724,676
PRIOR FILING DATE: 2000-11-28
PRIOR APPLICATION NUMBER: 1356776
PRIOR FILING DATE: 2000-06-15
PRIOR APPLICATION NUMBER: 135619
PRIOR FILING DATE: 2000-04-12
NUMBER OF SEQ ID NOS: 273
SOFTWARE: PatentIn version 3.0
SEQ ID NO 109
LENGTH: 213
TYPE: PRT
ORGANISM: Homo sapiens
US-09-771-161A-109

Query Match 2.2%; Score 7; DB 10; Length 213;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 41 STSGSGS 47

Db 187 STSGSGS 193

RESULT 8
US-10-125-540-300
Sequence 300, Application US/10125540
Publication No. US20030059875A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PZ14C1
CURRENT APPLICATION NUMBER: US/10/125,540
CURRENT FILING DATE: 2002-04-19
Prior Application removed - See file wrapper or Palm
NUMBER OF SEQ ID NOS: 646
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 300
LENGTH: 251
TYPE: PRT
ORGANISM: Homo sapiens
US-10-125-540-300

Query Match 2.2%; Score 7; DB 9; Length 251;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 167 KPTPTPP 173
Db 211 KPTPTPP 217

RESULT 9
US-09-764-870-300
Sequence 300, Application US/09764870
Patent No. US20020042386A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PZ14
CURRENT APPLICATION NUMBER: US/09/764,870
CURRENT FILING DATE: 2001-01-17
Prior application data removed - consult PALM or file wrapper
NUMBER OF SEQ ID NOS: 646
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 300
LENGTH: 251
TYPE: PRT
ORGANISM: Homo sapiens
US-09-764-870-300

Query Match 2.2%; Score 7; DB 10; Length 251;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 167 KPTPTPP 173
Db 211 KPTPTPP 217

RESULT 10
US-10-156-761-14239
Sequence 14239, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: MAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 14239
LENGTH: 343
TYPE: PRT
ORGANISM: Streptomyces avermitilis
US-10-156-761-14239

Query Match 2.2%; Score 7; DB 9; Length 343;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 215 GTATVAG 221
Db 63 GTATVAG 69

RESULT 11
US-09-943-671-4
Sequence 4, Application US/09943671
Patent No. US2002010673A1
GENERAL INFORMATION:
APPLICANT: Stahl, Andreas
APPLICANT: Hirsch, David J.
APPLICANT: Lodish, Harvey F.
TITLE OF INVENTION: Fatty Acid Transport Proteins
FILE REFERENCE: WHI97-21p3ME
CURRENT APPLICATION NUMBER: US/09/943,671
CURRENT FILING DATE: 2001-08-31
PRIOR APPLICATION NUMBER: 09/232,191
PRIOR FILING DATE: 1999-10-14
PRIOR APPLICATION NUMBER: 60/093,491
PRIOR FILING DATE: 1998-07-20
PRIOR APPLICATION NUMBER: 60/110,941
PRIOR FILING DATE: 1998-12-04
NUMBER OF SEQ ID NOS: 37
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO 4
LENGTH: 356
TYPE: PRT
ORGANISM: Saccharomyces cerevisiae
US-09-943-671-4

Query Match 2.2%; Score 7; DB 10; Length 356;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 186 TEAPFAT 192
Db 145 TEAPFAT 151

RESULT 12
US-10-156-761-8098
Sequence 8098, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMIURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761

CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 8098
LENGTH: 375
TYPE: PRT
ORGANISM: Streptomyces avermitilis
US-10-156-761-8098

Query Match 2.2%; Score 7; DB 9; Length 375;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 215 GTATVAG 221
Db 176 GTATVAG 182

RESULT 13
US-10-156-761-13366
Sequence 13366, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMIURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 13366
LENGTH: 490
TYPE: PRT
ORGANISM: Streptomyces avermitilis
US-10-156-761-13366

Query Match 2.2%; Score 7; DB 9; Length 490;
Best Local Similarity 100.0%; Pred. No. 3.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 179 APVAPV 185
Db 38 APVAPV 44

RESULT 14
US-10-081-872-212
Sequence 212, Application US/10081872
Publication No. US20030125534A1
GENERAL INFORMATION:
APPLICANT: Callen, Walter
APPLICANT: Richardson, Toby
APPLICANT: Frey, Gerhard
APPLICANT: Short, Jay M.
APPLICANT: Machur, Eric J.
APPLICANT: Gray, Kevin A.
APPLICANT: Kerovuo, Janne S.
APPLICANT: Slupeska, Malgorzata
TITLE OF INVENTION: ENZYMES HAVING ALPHA AMYLASE ACTIVITY
FILE REFERENCE: 09010-108001

CURRENT APPLICATION NUMBER: US/10/081,872
CURRENT FILING DATE: 2002-02-21
PRIOR APPLICATION NUMBER: US 60/270,495
PRIOR FILING DATE: 2001-02-21
PRIOR APPLICATION NUMBER: US 60/270,496
PRIOR FILING DATE: 2001-02-21
PRIOR APPLICATION NUMBER: US 60/291,122
PRIOR FILING DATE: 2001-05-14
NUMBER OF SEQ ID NOS: 321
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 212
LENGTH: 496
TYPE: PRT
ORGANISM: Environmental
US-10-081-872-212

Query Match 2.2%; Score 7; DB 9; Length 496;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 99 NINLSY 105
Db 167 NINLSY 173

RESULT 15
US-09-771-161A-200
Sequence 200, Application US/09771161A
Patent No. US20020110811A1
GENERAL INFORMATION:
APPLICANT: LEVINE, et al.
TITLE OF INVENTION: VARIANTS OF PROTEIN KINASES
FILE REFERENCE: 802620-2005.1
CURRENT FILING DATE: 2001-01-26
PRIOR APPLICATION NUMBER: US/09/771,161A
CURRENT FILING DATE: 2000-11-28
PRIOR APPLICATION NUMBER: 136776
PRIOR FILING DATE: 2000-06-15
PRIOR APPLICATION NUMBER: 135619
PRIOR FILING DATE: 2000-04-12
NUMBER OF SEQ ID NOS: 273
SOFTWARE: PatentIn version 3.0
SEQ ID NO 200
LENGTH: 505
TYPE: PRT
ORGANISM: Homo sapiens
US-09-771-161A-200

Query Match 2.2%; Score 7; DB 10; Length 505;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 41 STSGSGS 47
Db 187 STSGSGS 193

RESULT 16
US-09-903-068-8
Sequence 8, Application US/09903068
Patent No. US20020123139A1
GENERAL INFORMATION:
APPLICANT: Miyazono, Kohel; DiJke, Peter Ten;
Franzen, Petra; Yamashita, Hidetoshi; Heldin, Carl-Henrik
TITLE OF INVENTION: Activin Receptor-Like Kinase, Proteins
Having Serine Threonine Kinase Domains And Their Use
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Felfe & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York

ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage
COMPUTER: IBM
OPERATING SYSTEM: PC-DOS
SOFTWARE: wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/903,068
FILING DATE: 11-Jul-2001
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/679,187
FILING DATE: <unknown>
APPLICATION NUMBER: PCT/GB93/02367
FILING DATE: 17-No. US20020123139A1ember-1993
APPLICATION NUMBER: 9224057.1
FILING DATE: 17-No. US20020123139A1ember-1992
APPLICATION NUMBER: 9304677.9
FILING DATE: 8-March-1993
APPLICATION NUMBER: 9304680.3
FILING DATE: 8-March-1993
APPLICATION NUMBER: 9311047.6
FILING DATE: 28-May-1993
APPLICATION NUMBER: 9313763.6
FILING DATE: 2-July-1993
APPLICATION NUMBER: 9136099.2
FILING DATE: 3-August-1993
APPLICATION NUMBER: 9321344.5
FILING DATE: 15-October-1993
ATTORNEY/AGENT INFORMATION:
NAME: Kohel, Vineet
REGISTRATION NUMBER: 37,003
REFERENCE/DOCKET NUMBER: LUD 5298
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884

INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 505 amino acids
TYPE: amino acid
TOPOLOGY: linear

MOLECULE TYPE: Protein
SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-09-903-068-8

Query Match 2.2%; Score 7; DB 10; Length 505;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 41 STSGSGS 47
Db 187 STSGSGS 193

RESULT 17
US-09-903-068-16
Sequence 16, Application US/09903068
Patent No. US20020123139A1
GENERAL INFORMATION:
APPLICANT: Miyazono, Kohel; DiJke, Peter Ten;
Franzen, Petra; Yamashita, Hidetoshi; Heldin, Carl-Henrik
TITLE OF INVENTION: Activin Receptor-Like Kinase, Proteins
Having Serine Threonine Kinase Domains And Their Use
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Felfe & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage
COMPUTER: IBM
OPERATING SYSTEM: PC-DOS

SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/903.068
FILING DATE: 11-Jul-2001
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/679,187
FILING DATE: <Unknown>
APPLICATION NUMBER: PCT/GB93/02367
FILING DATE: 17-No. US20020123139A1ember-1993
APPLICATION NUMBER: 9224057.1
FILING DATE: 17-No. US20020123139A1ember-1992
APPLICATION NUMBER: 9304677.9
FILING DATE: 8-March-1993
APPLICATION NUMBER: 9304680.3
FILING DATE: 8-March-1993
APPLICATION NUMBER: 9311047.6
FILING DATE: 28-May-1993
APPLICATION NUMBER: 9313763.6
FILING DATE: 2-July-1993
APPLICATION NUMBER: 9136099.2
FILING DATE: 3-August-1993
APPLICATION NUMBER: 9321344.5
FILING DATE: 15-October-1993
ATTORNEY/AGENT INFORMATION:
NAME: Kohlei, Vineet
REGISTRATION NUMBER: 37,003
REFERENCE/DOCKET NUMBER: LUD 5298
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 638-3884
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 505 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 16:
US-09-903-068-16

Query Match 2.2%; Score 7; DB 10; Length 505;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 41 STSGSGS 47
Db 187 STSGSGS 193

RESULT 18
US-09-874-628-8
Sequence 8, Application US/09874628
Patent No. US20020137133A1
GENERAL INFORMATION:
APPLICANT: WOZNEY, John
CELESTE, Anthony J.
THIRS, R. Scott
YAMAJI, No. US20020137133A1
TITLE OF INVENTION: RECEPTOR PROTEINS
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute Inc.- Legal Affairs
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: MA
COUNTRY: USA
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/874,628

FILING DATE: 05-Jun-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/123,934
FILING DATE: 17-SEP-1993
ATTORNEY/AGENT INFORMATION:
NAME: LAZAR, Steven R
REGISTRATION NUMBER: 32,618
REFERENCE/DOCKET NUMBER: 5203
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617 876 1170
TELEFAX: 617 876 5851
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 505 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-09-874-628-8

Query Match 2.2%; Score 7; DB 10; Length 505;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 41 STSGSGS 47
Db 187 STSGSGS 193

RESULT 19
US-09-738-626-5890
Sequence 5890, Application US/09738626
Publication No. US20020197605A1
GENERAL INFORMATION:
APPLICANT: NAKAGAWA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: OCHIAI, KEIKO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, NAOKO
APPLICANT: SENOH, AKIHIRO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738,626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: Patentin ver. 3.0
SEQ ID NO 5890
LENGTH: 806
TYPE: PRT
ORGANISM: Corynebacterium glutamicum
US-09-738-626-5890

Query Match 2.2%; Score 7; DB 9; Length 806;
Best Local Similarity 100.0%; Pred. No. 5.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 230 WFSGRDG 236
Db 631 WFSGRDG 637

RESULT 20

US-09-978-303-25
Sequence 25, Application US/09978303
Publication No. US20030049728A1
GENERAL INFORMATION:
APPLICANT: Julie, David J.
APPLICANT: Caterina, Michael J.
APPLICANT: Blake, Anthony J.
TITLE OF INVENTION: Nucleic acid sequences encoding
TITLE OF INVENTION: capsacin receptor and capsacin receptor-related
FILE REFERENCE: UCA1084CON
CURRENT APPLICATION NUMBER: US/09/978,303
CURRENT FILING DATE: 2001-10-15
PRIOR APPLICATION NUMBER: 09/235,451
PRIOR FILING DATE: 1999-01-22
PRIOR APPLICATION NUMBER: 60/072,151
PRIOR FILING DATE: 1998-01-22
PRIOR APPLICATION NUMBER: 08/915,461
PRIOR FILING DATE: 1997-08-20
NUMBER OF SEQ ID NOS: 48
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 25
LENGTH: 843
TYPE: PRT
ORGANISM: chicken
US-09-978-303-25

Query Match 2.2%; Score 7; DB 9; Length 843;
Best Local Similarity 100.0%; Pred. No. 5.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 80 TVSKIAQ 86
Db 692 TVSKIAQ 698

RESULT 21
US-10-278-173-128
Sequence 128, Application US/10278173
Publication No. US20030061637A1
GENERAL INFORMATION:
APPLICANT: Jiang, Cai-Zhong
APPLICANT: Broun, Pierre
APPLICANT: Riechmann, Jose-Luis
APPLICANT: Pineda, Omatra
APPLICANT: Zhang, James
APPLICANT: Yu, Guo-Liang
APPLICANT: Pilgrim, Marsha
APPLICANT: Keddle, James
APPLICANT: Heard, Jacqueline
APPLICANT: Reuber, Lynne
APPLICANT: Ratcliffe, Oliver
APPLICANT: Adam, Luc
APPLICANT: Samaha, Raymond
TITLE OF INVENTION: POLYNUCLEOTIDES FOR ROOT TRAIT ALTERATION
FILE REFERENCE: MBI-009
CURRENT APPLICATION NUMBER: US/10/278,173
CURRENT FILING DATE: 2002-10-21
PRIOR APPLICATION NUMBER: US/09/533,392
PRIOR FILING DATE: 2000-03-22
PRIOR APPLICATION NUMBER: 60/125,814
PRIOR FILING DATE: 1999-03-23
NUMBER OF SEQ ID NOS: 177
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 128
LENGTH: 1336
TYPE: PRT
ORGANISM: Arabidopsis thaliana
FEATURE:
OTHER INFORMATION: G387
US-10-278-173-128

Query Match 2.2%; Score 7; DB 9; Length 1336;

Best Local Similarity 100.0%; Pred. No. 8.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 61 QVITDSQ 67
Db 779 QVITDSQ 785

RESULT 22
US-09-934-455-22
Sequence 22, Application US/09934455
Publication No. US20030121070A1
GENERAL INFORMATION:
APPLICANT: Adam, Luc
APPLICANT: Creelman, Robert
APPLICANT: Dubell, Arnold
APPLICANT: Heard, Jacqueline
APPLICANT: Jiang, Cai-Zhong
APPLICANT: Keddle, James
APPLICANT: Pilgrim, Marsha
APPLICANT: Ratcliffe, Oliver
APPLICANT: Reuber, Lynne
APPLICANT: Riechmann, Jose Luis
APPLICANT: Yu, Guo-Liang
APPLICANT: Pineda, Omatra
TITLE OF INVENTION: Genes for Modifying Plant Traits IV
FILE REFERENCE: MBI-0025
CURRENT APPLICATION NUMBER: US/09/934,455
CURRENT FILING DATE: 2001-08-22
PRIOR APPLICATION NUMBER: 60/227439
PRIOR FILING DATE: 2000-08-22
PRIOR APPLICATION NUMBER: MBI-0022
PRIOR FILING DATE: 2001-11-16
PRIOR APPLICATION NUMBER: MBI-0023
PRIOR FILING DATE: 2001-04-17
NUMBER OF SEQ ID NOS: 516
SOFTWARE: PatentIn version 3.1
SEQ ID NO 22
LENGTH: 1336
TYPE: PRT
ORGANISM: Arabidopsis thaliana
US-09-934-455-22

Query Match 2.2%; Score 7; DB 9; Length 1336;
Best Local Similarity 100.0%; Pred. No. 8.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 61 QVITDSQ 67
Db 779 QVITDSQ 785

RESULT 23
US-09-815-242-12713
Sequence 12713, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlisen, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA-011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848

;; PRIOR FILING DATE: 2000-05-23
;; PRIOR APPLICATION NUMBER: 60/207,727
;; PRIOR FILING DATE: 2000-05-26
;; PRIOR APPLICATION NUMBER: 60/242,578
;; PRIOR FILING DATE: 2000-10-23
;; PRIOR APPLICATION NUMBER: 60/253,625
;; PRIOR FILING DATE: 2000-11-27
;; PRIOR APPLICATION NUMBER: 60/257,931
;; PRIOR FILING DATE: 2000-12-22
;; PRIOR APPLICATION NUMBER: 60/269,308
;; PRIOR FILING DATE: 2001-02-16
;; NUMBER OF SEQ ID NOS: 14110
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 12713
;; LENGTH: 2344
;; TYPE: PRT
;; ORGANISM: Staphylococcus aureus
US-09-815-242-12713

Query Match 2.2%; Score 7; DB 10; Length 2344;
Best Local Similarity 100.0%; Pred. No. 1.4e+03;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 STSGSGS 47
DB 2088 STSGSGS 2094

RESULT 24
US-09-870-759-114
;; Sequence 114, Application US/09870759
;; Patent No. US20020177551A1
;; GENERAL INFORMATION:
;; APPLICANT: TERMAN, David S
;; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT OF NEOPLASTIC DISEASE
;; FILE REFERENCE: 870759
;; CURRENT APPLICATION NUMBER: US/09/870,759
;; CURRENT FILING DATE: 2002-01-14
;; PRIOR APPLICATION NUMBER: US 60/208,128
;; PRIOR FILING DATE: 2000-05-30
;; NUMBER OF SEQ ID NOS: 166
;; SOFTWARE: PatentIn version 3.1
;; SEQ ID NO 114
;; LENGTH: 2796
;; TYPE: PRT
;; ORGANISM: Mycobacterium bovis
US-09-870-759-114

Query Match 2.2%; Score 7; DB 9; Length 2796;
Best Local Similarity 100.0%; Pred. No. 1.6e+03;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 272 IHKDAQ 278
DB 2088 IHKDAQ 2094

RESULT 25
US-09-712-363-246
;; Sequence 246, Application US/09712363
;; Patent No. US20020164588A1
;; GENERAL INFORMATION:
;; APPLICANT: Eisenberg, David
;; APPLICANT: Rotstein, Sergio H.
;; APPLICANT: Marcotte, Edward M.
;; TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND
;; FILE REFERENCE: 07419-032001
;; CURRENT APPLICATION NUMBER: US/09/712,363
;; CURRENT FILING DATE: 2000-11-13
;; PRIOR APPLICATION NUMBER: PCT/US00/02246
;; PRIOR FILING DATE: 2000-01-28
;; PRIOR APPLICATION NUMBER: 60/179,531

;; PRIOR FILING DATE: 2000-02-01
;; PRIOR APPLICATION NUMBER: 60/117,844
;; PRIOR FILING DATE: 1999-01-29
;; PRIOR APPLICATION NUMBER: 60/118,206
;; PRIOR FILING DATE: 1999-02-01
;; PRIOR APPLICATION NUMBER: 60/126,593
;; PRIOR FILING DATE: 1999-03-26
;; PRIOR APPLICATION NUMBER: 60/134,093
;; PRIOR FILING DATE: 1999-05-14
;; PRIOR APPLICATION NUMBER: 60/134,092
;; PRIOR FILING DATE: 1999-05-14
;; PRIOR APPLICATION NUMBER: 60/165,124
;; PRIOR FILING DATE: 1999-11-12
;; PRIOR APPLICATION NUMBER: 60/165,086
;; PRIOR FILING DATE: 1999-11-12
;; NUMBER OF SEQ ID NOS: 292
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 246
;; LENGTH: 3069
;; TYPE: PRT
;; ORGANISM: Mycobacterium tuberculosis
US-09-712-363-246

Query Match 2.2%; Score 7; DB 9; Length 3069;
Best Local Similarity 100.0%; Pred. No. 1.8e+03;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 272 IHKDAQ 278
DB 2206 IHKDAQ 2212

RESULT 26
US-09-956-086-7
;; Sequence 7, Application US/09956086
;; Patent No. US20020155498A1
;; GENERAL INFORMATION:
;; APPLICANT: FILPULA, DAVID
;; WANG, MAOLIANG
;; SHORR, ROBERT
;; WHITLOW, MARC
;; LEE, LILSYNG S.
;; TITLE OF INVENTION: SINGLE-CHAIN ANTIGEN-BINDING PROTEINS
;; CAPABLE OF GLYCOSYLATION, PRODUCTION AND USES THEREOF
;; NUMBER OF SEQUENCES: 33
;; CORRESPONDENCE ADDRESSES:
;; ADDRESSEE: STERN, KESSLER, GOLDSTEIN & FOX P.L.L.C.
;; STREET: 1100 NEW YORK AVE., NW, SUITE 600
;; CITY: WASHINGTON
;; STATE: DC
;; COUNTRY: USA
;; ZIP: 20005
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/956,086
;; FILING DATE: 20-Sep-2001
;; CLASSIFICATION: <Unknown>
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 09/069,821
;; FILING DATE: <Unknown>
;; APPLICATION NUMBER: US 60/063,074
;; FILING DATE: 27-OCT-1997
;; APPLICATION NUMBER: US 60/050,472
;; FILING DATE: 23-JUN-1997
;; APPLICATION NUMBER: US 60/044,449
;; FILING DATE: 30-APR-1997
;; ATTORNEY/AGENT INFORMATION:
;; NAME: KIM, JUDITH U.
;; REGISTRATION NUMBER: 40,679

```

; REFERENCE/DOCKET NUMBER: 0977.2280003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)371-2600
; TELEFAX: (202)371-2540
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: No. US20020155498A1 Relevant
; MOLECULE TYPE: Peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-09-956-086-7

Query Match          1.9%; Score 6; DB 9; Length 14;
Best Local Similarity 100.0%; Pred.No.1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 STSGSG 46
Db 2 STSGSG 7

RESULT 27
US-09-999-745-58
; Sequence 58, Application US/09999745
; Patent No. US20020157120A1
; GENERAL INFORMATION:
; APPLICANT: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
; APPLICANT: Tsien, Roger Y.
; TITLE OF INVENTION: CIRCULARLY PERMUTED FLUORESCENT PROTEIN INDICATORS
; FILE REFERENCE: REGEN1470-1
; CURRENT APPLICATION NUMBER: US/09/999,745
; CURRENT FILING DATE: 2001-10-23
; PRIOR APPLICATION NUMBER: 09/316,920
; PRIOR FILING DATE: 1999-05-21
; NUMBER OF SEQ ID NOS: 67
; SOFTWARE: Patent version 3.0
; SEQ ID NO 58
; LENGTH: 14
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Linker moiety
US-09-999-745-58

Query Match          1.9%; Score 6; DB 9; Length 14;
Best Local Similarity 100.0%; Pred.No.1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 STSGSG 46
Db 2 STSGSG 7

RESULT 28
US-09-999-745-60
; Sequence 60, Application US/09999745
; Patent No. US20020157120A1
; GENERAL INFORMATION:
; APPLICANT: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
; APPLICANT: Tsien, Roger Y.
; TITLE OF INVENTION: CIRCULARLY PERMUTED FLUORESCENT PROTEIN INDICATORS
; FILE REFERENCE: REGEN1470-1
; CURRENT APPLICATION NUMBER: US/09/999,745
; CURRENT FILING DATE: 2001-10-23
; PRIOR APPLICATION NUMBER: 09/316,920
; PRIOR FILING DATE: 1999-05-21
; NUMBER OF SEQ ID NOS: 67
; SOFTWARE: Patent version 3.0
; SEQ ID NO 60
```

```

; LENGTH: 14
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Linker moiety
US-09-999-745-60

Query Match          1.9%; Score 6; DB 9; Length 14;
Best Local Similarity 100.0%; Pred.No.1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 STSGSG 46
Db 2 STSGSG 7

RESULT 29
US-09-956-087-7
; Sequence 7, Application US/09956087
; Patent No. US20020161201A1
; GENERAL INFORMATION:
; APPLICANT: FILIPULA, DAVID
; APPLICANT: WANG, MAOLIANG
; APPLICANT: SHORR, ROBERT
; APPLICANT: WHITLOW, MARC
; APPLICANT: LEE, LHSYNG S.
; TITLE OF INVENTION: SINGLE-CHAIN ANTIGEN-BINDING PROTEINS
; CAPABLE OF GLYCOSYLATION, PRODUCTION AND USES THEREOF
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERN, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVE., NW, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/956,087
; FILING DATE: 20-Sep-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/069,821
; FILING DATE: 1998-04-30
; APPLICATION NUMBER: US 60/063,074
; FILING DATE: 27-OCT-1997
; APPLICATION NUMBER: US 60/050,472
; FILING DATE: 23-JUN-1997
; APPLICATION NUMBER: US 60/044,449
; FILING DATE: 30-APR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: KIM, JUDITH U
; REGISTRATION NUMBER: 40,679
; REFERENCE/DOCKET NUMBER: 0977.2280003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)371-2600
; TELEFAX: (202)371-2540
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: not relevant
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-09-956-087-7

Query Match          1.9%; Score 6; DB 9; Length 14;
Best Local Similarity 100.0%; Pred.No.1.3e+02;
```

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 STSGSG 46
|||||
Db 2 STSGSG 7

RESULT 30
US-09-554-000-42
; Sequence 42, Application US/09554000
; Patent No. US20020165364A1
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Miyawaki, Atsushi
; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
; TITLE OF INVENTION: DETECTION OF ANALYTES
; FILE REFERENCE: 07257/042001
; CURRENT APPLICATION NUMBER: US/09/554,000
; CURRENT FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: 08/818,252
; PRIOR FILING DATE: 1997-03-14
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: FastSeq For Windows Version 4.0
; SEQ ID NO 42
; LENGTH: 14
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Linker moiety
US-09-554-000-42

Query Match 1.9%; Score 6; DB 9; Length 14;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 STSGSG 46
|||||
Db 2 STSGSG 7

RESULT 31
US-09-554-000-44
; Sequence 44, Application US/09554000
; Patent No. US20020165364A1
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Miyawaki, Atsushi
; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
; TITLE OF INVENTION: DETECTION OF ANALYTES
; FILE REFERENCE: 07257/042001
; CURRENT APPLICATION NUMBER: US/09/554,000
; CURRENT FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: 08/818,252
; PRIOR FILING DATE: 1997-03-14
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: FastSeq For Windows Version 4.0
; SEQ ID NO 44
; LENGTH: 14
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Linker moiety
US-09-554-000-44

Query Match 1.9%; Score 6; DB 9; Length 14;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 STSGSG 46
|||||
Db 2 STSGSG 7

RESULT 32
US-09-792-793A-4
; Sequence 4, Application US/09792793A
; Patent No. US20020168370A1
; GENERAL INFORMATION:
; APPLICANT: McDonald, John R.
; APPLICANT: Cogging, Philip
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR TREATING SECONDARY TISSUE DAMAGE AND
; TITLE OF INVENTION: OTHER INFLAMMATORY CONDITIONS AND DISORDERS
; FILE REFERENCE: 25020-601D
; CURRENT APPLICATION NUMBER: US/09/792,793A
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 93
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 14
; TYPE: PRT
; ORGANISM: homo sapien
; FEATURE:
; OTHER INFORMATION: peptide linker for use in conjugates
US-09-792-793A-4

Query Match 1.9%; Score 6; DB 9; Length 14;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 STSGSG 46
|||||
Db 2 STSGSG 7

RESULT 33
US-09-792-793A-6
; Sequence 6, Application US/09792793A
; Patent No. US20020168370A1
; GENERAL INFORMATION:
; APPLICANT: McDonald, John R.
; APPLICANT: Cogging, Philip
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR TREATING SECONDARY TISSUE DAMAGE AND
; TITLE OF INVENTION: OTHER INFLAMMATORY CONDITIONS AND DISORDERS
; FILE REFERENCE: 25020-601D
; CURRENT APPLICATION NUMBER: US/09/792,793A
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 93
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 14
; TYPE: PRT
; ORGANISM: homo sapien
; FEATURE:
; OTHER INFORMATION: peptide linker for use in conjugates
US-09-792-793A-6

Query Match 1.9%; Score 6; DB 9; Length 14;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 STSGSG 46
|||||
Db 2 STSGSG 7

RESULT 34
US-10-158-238-9
; Sequence 9, Application US/10158238
; Publication No. US20030040604A1
; GENERAL INFORMATION:
; APPLICANT: Immunex Corporation
; APPLICANT: Marken, John S.
; APPLICANT: Marken, John S.
; TITLE OF INVENTION: STGEC-12 POLYPEPTIDES, POLYNUCLEOTIDES, AND METHODS OF USE THERE
; FILE REFERENCE: 3290-A
; CURRENT APPLICATION NUMBER: US/10/158,238

Wed Jul 9 10:00:58 2003

us-10-018-706-2.cligo.rapb

Page 18

CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: US 60/294,199
PRIOR FILING DATE: 2001-05-29
NUMBER OF SEQ ID NOS: 30
SOFTWARE: PatentIn version 3.1
SEQ ID NO 9
LENGTH: 14
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Peptide linker sequence
US-10-158-238-9

Query Match 1.9%; Score 6; DB 9; Length 14;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 41 STSGSG 46
Db 2 STSGSG 7

RESULT 35
US-10-247-279-3
Sequence 3, Application US/10247279
Publication No. US20030086928A1
GENERAL INFORMATION:
APPLICANT: Montminy, Marc
TITLE OF INVENTION: COMPOSITIONS ASSOCIATED WITH COMPLEX
FILE REFERENCE: SALKINS.036CP1
CURRENT FILING DATE: 2002-09-17
PRIOR APPLICATION NUMBER: 60/190,705
PRIOR FILING DATE: 2000-03-17
PRIOR APPLICATION NUMBER: PCTUS 01/08946
PRIOR FILING DATE: 2001-03-19
NUMBER OF SEQ ID NOS: 13
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 14
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Linker
US-10-247-279-3

Query Match 1.9%; Score 6; DB 9; Length 14;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 41 STSGSG 46
Db 2 STSGSG 7

RESULT 36
US-10-052-942-6
Sequence 6, Application US/10052942
Publication No. US20030104402A1
GENERAL INFORMATION:
APPLICANT: Zauderer, Maurice
APPLICANT: Smith, Ernest
APPLICANT: Wei, Chungwen
TITLE OF INVENTION: Methods of Producing or Identifying Intrabodies in Eukaryotic Cell
FILE REFERENCE: 1821.0090004
CURRENT APPLICATION NUMBER: US/10/052,942
CURRENT FILING DATE: 2002-01-23
PRIOR APPLICATION NUMBER: 60/298,095
PRIOR FILING DATE: 2001-06-15
PRIOR APPLICATION NUMBER: 60/271,422
PRIOR FILING DATE: 2001-02-27

PRIOR APPLICATION NUMBER: 60/263,200
PRIOR FILING DATE: 2001-01-24
PRIOR APPLICATION NUMBER: 60/263,225
PRIOR FILING DATE: 2001-01-23
NUMBER OF SEQ ID NOS: 154
SOFTWARE: PatentIn version 3.0
SEQ ID NO 6
LENGTH: 14
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Linker
US-10-052-942-6

Query Match 1.9%; Score 6; DB 9; Length 14;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 41 STSGSG 46
Db 2 STSGSG 7

RESULT 37
US-09-883-777-13
Sequence 13, Application US/09883777
Patent No. US20020110853A1
GENERAL INFORMATION:
APPLICANT: Wiley, Steven R.
TITLE OF INVENTION: TWEAK RECEPTOR
FILE REFERENCE: 2868-C
CURRENT APPLICATION NUMBER: US/09/883,777
CURRENT FILING DATE: 2001-06-18
PRIOR APPLICATION NUMBER: US 60/172,878
PRIOR FILING DATE: 1998-12-20
PRIOR APPLICATION NUMBER: US 60/203,347
PRIOR FILING DATE: 2000-05-10
PRIOR APPLICATION NUMBER: PCT/US00/34755
PRIOR FILING DATE: 2000-12-19
PRIOR APPLICATION NUMBER: US 09/742,454
PRIOR FILING DATE: 2000-12-19
NUMBER OF SEQ ID NOS: 16
SOFTWARE: PatentIn version 3.1
SEQ ID NO 13
LENGTH: 14
TYPE: PRT
ORGANISM: Artificial sequence
FEATURE:
OTHER INFORMATION: Linker Moleley
US-09-883-777-13

Query Match 1.9%; Score 6; DB 10; Length 14;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 41 STSGSG 46
Db 2 STSGSG 7

RESULT 38
US-09-987-456-11
Sequence 11, Application US/09987456
Patent No. US20020123057A1
GENERAL INFORMATION:
APPLICANT: University of Rochester
APPLICANT: Zauderer, Maurice
APPLICANT: Ernest S. Smith
TITLE OF INVENTION: In Vitro Methods of Producing and Selecting
FILE REFERENCE: 1821.0070004
CURRENT APPLICATION NUMBER: US/09/987,456
CURRENT FILING DATE: 2001-11-14

; PRIOR APPLICATION NUMBER: 60/271,424
; PRIOR FILING DATE: 2001-02-27
; PRIOR APPLICATION NUMBER: 60/262,067
; PRIOR FILING DATE: 2001-01-18
; PRIOR APPLICATION NUMBER: 60/298,087
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: 60/249,268
; PRIOR FILING DATE: 2000-11-17
; NUMBER OF SEQ ID NOS: 147
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 11
; LENGTH: 14
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: linker peptide
US-09-987-456-11

Query Match 1.9%; Score 6; DB 10; Length 14;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 41 STSGSG 46
|||
Db 2 STSGSG 7

RESULT 39
US-09-883-777-14
; Sequence 14, Application US/09883777
; Patent No. US20020110853A1
; GENERAL INFORMATION:
; APPLICANT: Wiley, Steven R.
; TITLE OF INVENTION: TWEAK RECEPTOR
; FILE REFERENCE: 2968-C
; CURRENT APPLICATION NUMBER: US/09/883,777
; CURRENT FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: US 60/172,878
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: US 60/203,347
; PRIOR FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: PCT/US00/34755
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: US 09/742,454
; PRIOR FILING DATE: 2000-12-19
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 14
; LENGTH: 16
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Linker Moiety
US-09-883-777-14

Query Match 1.9%; Score 6; DB 10; Length 16;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 41 STSGSG 46
|||
Db 2 STSGSG 7

RESULT 40
US-09-883-777-15
; Sequence 15, Application US/09883777
; Patent No. US20020110853A1
; GENERAL INFORMATION:
; APPLICANT: Wiley, Steven R.
; TITLE OF INVENTION: TWEAK RECEPTOR
; FILE REFERENCE: 2968-C
; CURRENT APPLICATION NUMBER: US/09/883,777

; CURRENT FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: US 60/172,878
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: US 60/203,347
; PRIOR FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: PCT/US00/34755
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: US 09/742,454
; PRIOR FILING DATE: 2000-12-19
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 15
; LENGTH: 16
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Linker Moiety
US-09-883-777-15

Query Match 1.9%; Score 6; DB 10; Length 16;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 41 STSGSG 46
|||
Db 2 STSGSG 7

Search completed: July 8, 2003, 11:30:23
Job time : 74 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 8, 2003, 11:16:18 ; Search time 40 Seconds
(without alignments)
773.882 Million cell updates/sec

Title: US-10-018-706-2

Perfect score: 322

Sequence: 1 MVTIAINSQNKPIKRLGH.....LFEFRISRGVYDPLTLK 322

Scoring table: OLIGO

Gapop 60.0 , Gapept 60.0

Searched: 283224 seqs, 96134422 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database :

1: PIR_73:*
2: PIR1:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8	2.5	164	2	A69865 conserved hypochet
2	8	2.5	227	2	H84135 hypochetrical prote
3	8	2.5	244	2	S50385 hypochetrical prote
4	8	2.5	306	2	T52340 cell wall-plasma m
5	8	2.5	376	2	S71558 probable cell wall
6	8	2.5	968	2	S63211 cytoskeleton assem
7	7	2.2	101	2	H86885 50S ribosomal prot
8	7	2.2	115	2	S69849 hypochetrical prote
9	7	2.2	139	2	T11639 ribosomal protein
10	7	2.2	141	2	T29781 hypochetrical prote
11	7	2.2	161	2	T31531 hypochetrical prote
12	7	2.2	177	2	T27153 hypochetrical prote
13	7	2.2	189	2	A10213 probable membrane
14	7	2.2	198	2	UC6024 kinB sporulation s
15	7	2.2	208	2	G86352 dedd protein - Bsc
16	7	2.2	211	1	KMECD1 probable membrane
17	7	2.2	225	2	A69305 conserved hypochet
18	7	2.2	227	2	S26440 hypochetrical prote
19	7	2.2	225	2	A69305 conserved hypochet
20	7	2.2	241	2	H82435 conserved hypochet
21	7	2.2	249	2	C84340 ABC transport prot
22	7	2.2	259	2	C70151 hypochetrical prote
23	7	2.2	269	2	F89455 protein F55A4.4 [i
24	7	2.2	283	2	S57134 hypochetrical prote
25	7	2.2	284	2	H98073 hypochetrical prote
26	7	2.2	285	2	D69184 UTP-glucose-1-phos
27	7	2.2	288	2	JH0204 hypochetrical 30.5K
28	7	2.2	303	2	S28264 hydroxyproline-ric
29	7	2.2	309	2	AE3325 comL, competence 1

30	7	2.2	315	2	G84936
31	7	2.2	326	2	T26886
32	7	2.2	328	2	J00985
33	7	2.2	329	2	T40981
34	7	2.2	338	2	T40448
35	7	2.2	339	2	T34158
36	7	2.2	346	2	S19129
37	7	2.2	350	2	S22456
38	7	2.2	351	2	AC0302
39	7	2.2	355	2	G71958
40	7	2.2	367	2	H83202
41	7	2.2	368	2	UN0777
42	7	2.2	368	2	T26269
43	7	2.2	369	2	S20500
44	7	2.2	377	2	G97314
45	7	2.2	381	2	T27806
46	7	2.2	387	1	RSB14E
47	7	2.2	387	2	AD2335
48	7	2.2	387	2	T31748
49	7	2.2	391	2	E82668
50	7	2.2	409	1	S32905
51	7	2.2	409	2	T35155
52	7	2.2	413	2	AH2743
53	7	2.2	414	2	T13436
54	7	2.2	416	2	G97524
55	7	2.2	421	2	H84807
56	7	2.2	422	2	T09742
57	7	2.2	424	2	P95330
58	7	2.2	429	2	AE2501
59	7	2.2	433	2	T46528
60	7	2.2	449	2	S67819
61	7	2.2	471	2	S30585
62	7	2.2	476	2	T80182
63	7	2.2	478	1	S73920
64	7	2.2	487	2	T80183
65	7	2.2	499	2	A11107
66	7	2.2	505	2	T13859
67	7	2.2	505	2	T53417
68	7	2.2	513	2	T40998
69	7	2.2	517	2	T27927
70	7	2.2	532	2	S18643
71	7	2.2	542	2	AC0414
72	7	2.2	542	2	T29707
73	7	2.2	551	1	A55582
74	7	2.2	580	1	MMBEM5
75	7	2.2	603	2	F72237
76	7	2.2	618	1	PRECT4
77	7	2.2	618	2	H90937
78	7	2.2	618	2	D85786
79	7	2.2	623	2	S45899
80	7	2.2	634	2	T18711
81	7	2.2	634	2	T18702
82	7	2.2	635	2	T06537
83	7	2.2	652	2	E97857
84	7	2.2	729	2	B41670
85	7	2.2	731	2	T16524
86	7	2.2	772	2	T08826
87	7	2.2	781	2	T18693
88	7	2.2	814	2	T49207
89	7	2.2	836	2	T33222
90	7	2.2	865	2	A47282
91	7	2.2	873	2	A47283
92	7	2.2	886	2	S07132
93	7	2.2	887	2	AD2009
94	7	2.2	1047	2	A59246
95	7	2.2	1119	2	AB2239
96	7	2.2	1248	2	C89874
97	7	2.2	1305	2	AH3452
98	7	2.2	1336	2	T02736
99	7	2.2	1388	2	A53317
100	7	2.2	1428	2	T08852
101	7	2.2	1672	2	C81675
102	7	2.2	1965	2	T33216

kinase [imported]
hypochetrical prote
hydroxyproline-ric
probable quinone o
conserved hypochet
hypochetrical prote
proline-rich prote
hydroxyproline-ric
conserved hypochet
hypochetrical prote
poly(beta-d-mannur
poly(beta-d-mannur
hypochetrical prote
hydroxyproline-ric
alcohol dehydrogen
hypochetrical prote
ribosomal protein
ornithine acetyl t
hypochetrical prote
dihydrolipoamide S
serine proteinase
hypochetrical prote
conserved hypochet
hypochetrical prote
hypochetrical prote
probable phospholi
drought-induced pr
hypochetrical prote
hypochetrical prote
probable GDP-4-ket
Gumc protein - Xan
hypochetrical prote
activin type I rec
amidase homolog GO
activin type I rec
internalin B [limp
activin A receptor
type I serine-thre
hypochetrical prote
hypochetrical prote
maternal exupertant
probable exported
hypochetrical prote
cytochrome-c oxida
U125 protein - hum
conserved hypochet
proteinase IV (EC
proteinase IV [limp
proteinase IV, a s
probable membrane
hypochetrical prote
hypochetrical prote
ethylene receptor
cell surface antig
carbon-monoxide de
hypochetrical prote
hypochetrical prote
receptor kinase-II
hypochetrical prote
calcium-binding pr
calpoptin - fruit
hypochetrical prote
hypochetrical prote
HIRA protein - fru
hypochetrical prote
autolysin [importe
sensory transducto
probable SCARROW
collagen alpha 1(X
lustrin A - Califo
polymorphic membra
hypochetrical prote

103	2.2	2237	2	T21087	176	6	1.9	126	2	T40732	probable 50s ribos
104	2.2	2245	2	T21086	177	6	1.9	127	2	C84502	Bn/SpM-like transp
105	2.2	2271	2	F90073	178	6	1.9	128	2	T13112	protein gp26 - pha
106	2.2	2281	2	T28155	179	6	1.9	129	2	S08609	surface glycoprote
107	2.2	2796	2	UC4743	180	6	1.9	130	2	S28009	root-specific prote
108	2.2	3069	7	H70656	181	6	1.9	131	2	A11875	hypothetical prote
109	2.2	3076	2	A87058	182	6	1.9	132	2	C87469	hypothetical prote
110	1.9	27	2	A43768	183	6	1.9	133	2	S57882	T-cell receptor PS
111	1.9	33	2	A21211	184	6	1.9	135	2	T09033	hypothetical prote
112	1.9	40	2	S21044	185	6	1.9	135	2	AD1828	hypothetical prote
113	1.9	51	2	I51028	186	6	1.9	136	2	H83300	hypothetical prote
114	1.9	56	2	I10959	187	6	1.9	137	2	S35714	proline-rich prote
115	1.9	57	2	S12957	188	6	1.9	137	2	T22308	hypothetical prote
116	1.9	57	2	C71442	189	6	1.9	138	2	H87362	organic hydroperox
117	1.9	57	2	AH0423	190	6	1.9	138	2	T29558	hypothetical prote
118	1.9	60	2	S10210	191	6	1.9	139	2	A90454	conserved thioether
119	1.9	60	2	S47158	192	6	1.9	140	2	G70664	probable thioether
120	1.9	79	2	B83400	193	6	1.9	141	2	T09546	extensin like prot
121	1.9	80	2	PC4370	194	6	1.9	142	2	E86648	reductase [import
122	1.9	83	2	G82744	195	6	1.9	142	2	S51582	galactoside 2-alph
123	1.9	84	2	G87599	196	6	1.9	144	2	AF1780	hypothetical prote
124	1.9	84	2	T16647	197	6	1.9	144	2	AG1404	hypothetical prote
125	1.9	85	1	MOECBH	198	6	1.9	147	2	S42552	proline-rich prote
126	1.9	85	1	MOEBBH	199	6	1.9	147	2	S67304	probable membrane
127	1.9	85	1	S12749	200	6	1.9	148	2	T02305	hypothetical prote
128	1.9	85	2	A82248	201	6	1.9	148	2	G88642	protein C54B4.4 [1
129	1.9	85	2	AB0364	202	6	1.9	148	2	F71454	hypothetical prote
130	1.9	85	2	AB0364	203	6	1.9	149	2	D97218	probable membrane
131	1.9	85	2	AB0364	204	6	1.9	151	2	D69132	nucleoside-diphosp
132	1.9	85	2	G91039	205	6	1.9	151	2	G69345	nucleoside-diphosp
133	1.9	86	2	C41139	206	6	1.9	151	2	T38090	ACMPV orf91 - Bom
134	1.9	87	1	CCF26	207	6	1.9	154	2	T41831	probable bacteriof
135	1.9	87	2	G90954	208	6	1.9	154	2	G70819	hypothetical prote
136	1.9	87	2	T18152	209	6	1.9	155	2	T25980	hypothetical prote
137	1.9	87	2	D85803	210	6	1.9	155	2	E82843	ribosomal protein
138	1.9	89	2	G81106	211	6	1.9	156	2	RS8SH2	50S ribosomal prot
139	1.9	89	2	D82298	212	6	1.9	156	2	A84322	immediate-early pr
140	1.9	89	2	AB1241	213	6	1.9	157	2	EDX73	luteinizing hormon
141	1.9	89	2	AF1603	214	6	1.9	159	2	T15173	hypothetical prote
142	1.9	89	2	UC7825	215	6	1.9	159	2	T48837	hypothetical prote
143	1.9	95	2	A41872	216	6	1.9	159	2	S65994	YyC protein - Bac
144	1.9	95	2	G97232	217	6	1.9	159	2	E69531	molycoperoxin-gua
145	1.9	99	2	A58889	218	6	1.9	159	2	T48209	galactoside 2-alph
146	1.9	99	2	PC4431	219	6	1.9	159	2	S46494	hypothetical prote
147	1.9	100	2	T49636	220	6	1.9	160	2	A83110	hypothetical prote
148	1.9	100	2	S38868	221	6	1.9	160	2	B71176	hypothetical prote
149	1.9	101	2	H84116	222	6	1.9	161	2	A95049	acetyl-CoA carboxy
150	1.9	101	2	T41466	223	6	1.9	161	2	G97919	isocitrate lyase (
151	1.9	103	2	A81032	224	6	1.9	164	2	H81808	ompl6 protein - Br
152	1.9	103	2	B81975	225	6	1.9	164	2	T04115	peptidoglycan-asso
153	1.9	103	2	S72925	226	6	1.9	168	2	A40346	neurotensin / neur
154	1.9	103	2	S36545	227	6	1.9	168	2	AF3294	probable acetoin u
155	1.9	103	2	T36233	228	6	1.9	169	2	A28146	dihydrofolate redu
156	1.9	106	2	S08315	229	6	1.9	169	2	D82284	conserved hypotet
157	1.9	109	2	G71013	230	6	1.9	169	2	T18321	hypothetical prote
158	1.9	109	2	B2746	231	6	1.9	171	2	E87512	hypothetical prote
159	1.9	110	2	H86450	232	6	1.9	174	2	G81656	hypothetical prote
160	1.9	110	2	PC4432	233	6	1.9	174	2	H71498	hypothetical prote
161	1.9	113	2	D90171	234	6	1.9	174	2	E92900	hypothetical prote
162	1.9	114	2	B29029	235	6	1.9	175	2	C84254	probable membrane
163	1.9	114	2	F83035	236	6	1.9	176	2	AC0635	hypothetical prote
164	1.9	115	1	JN0541	237	6	1.9	176	2	B82478	hypothetical prote
165	1.9	115	2	E90767	238	6	1.9	176	2	T43308	hypothetical prote
166	1.9	115	2	C90970	239	6	1.9	178	2	C75059	hypothetical prote
167	1.9	115	2	B85717	240	6	1.9	178	2	D72507	ribonuclease H
168	1.9	115	2	G72568	241	6	1.9	179	2	G75462	probable membrane
169	1.9	119	2	A82506	242	6	1.9	179	2	A64846	hypothetical prote
170	1.9	121	2	S05720	243	6	1.9	179	2	E90805	hypothetical prote
171	1.9	121	2	H72507	244	6	1.9	179	2	B85665	gas-vesicle operon
172	1.9	123	2	C76522	245	6	1.9	181	2	T44972	probable 11pope
173	1.9	123	2	S78697	246	6	1.9	184	2	AE0071	dhhydridipicolinat
174	1.9	125	2	AC0753	247	6	1.9	185	2	S60778	peptidylprolyl iso
175	1.9	125	2	T51037	248	6	1.9	185	2	H81837	

249	6	1.9	185	2	H81103	peptidylprolyl iso	322	6	1.9	232	2	B81339	probable flagellar
250	6	1.9	185	2	C70550	hypothetical prote	323	6	1.9	233	1	C39142	mkB protein precu
251	6	1.9	186	2	C72782	hypothetical prote	324	6	1.9	233	2	T32500	hypothetical prote
252	6	1.9	187	1	B29830	toluene dioxygenas	325	6	1.9	234	2	AG1320	glycerol uptake fa
253	6	1.9	187	2	S76728	hypothetical prote	326	6	1.9	234	2	S60885	terric exochelin u
254	6	1.9	187	2	T50946	hypothetical prote	327	6	1.9	236	1	B65073	protein disulfide
255	6	1.9	188	2	S74560	hypothetical prote	328	6	1.9	236	2	E91099	protein disulfide
256	6	1.9	188	2	D87705	intracellular sept	329	6	1.9	236	2	A85945	hypothetical prote
257	6	1.9	190	2	S75069	K+-transporting AT	330	6	1.9	236	2	I40239	hypothetical prote
258	6	1.9	191	2	T04853	hypothetical prote	331	6	1.9	237	2	E84311	hypothetical prote
259	6	1.9	191	2	AH0186	probable lipoprote	332	6	1.9	239	2	D70854	hypothetical prote
260	6	1.9	192	2	F70005	hypothetical prote	333	6	1.9	241	2	B49202	cysteine-rich secr
261	6	1.9	192	2	A10383	probable lipoprote	334	6	1.9	241	2	T05040	nodulin-26-like pr
262	6	1.9	192	2	E96940	uncharacterized me	335	6	1.9	241	2	H72691	probable hexulose
263	6	1.9	194	2	J01095	hypothetical 21.8x	336	6	1.9	242	2	T45421	probable 2-hydroxy
264	6	1.9	194	2	E90223	rRNA adenine N-6-m	337	6	1.9	242	2	A82754	lipoprotein Xf0855
265	6	1.9	194	2	JC7282	flavodoxin FRK I	338	6	1.9	243	2	S44914	Zf68.7 protein -
266	6	1.9	195	2	E64755	YagZ protein - Esc	339	6	1.9	244	2	S73022	probable membrane
267	6	1.9	195	2	C90669	hypothetical prote	340	6	1.9	244	2	C97532	transcription regu
268	6	1.9	195	2	F85519	hypothetical prote	341	6	1.9	244	2	AC2751	hypothetical prote
269	6	1.9	195	2	A13398	probable capsule p	342	6	1.9	246	2	C84998	hypothetical prote
270	6	1.9	196	2	AC3300	conserved membrane-	343	6	1.9	248	2	T26461	hypothetical prote
271	6	1.9	196	2	I64222	conserved hypochet	344	6	1.9	248	2	T33230	hypothetical prote
272	6	1.9	197	2	C72545	probable 3-octapre	345	6	1.9	248	2	T02122	hypothetical prote
273	6	1.9	197	2	T36071	probable pantoate-	346	6	1.9	249	2	T25643	hypothetical prote
274	6	1.9	197	2	S53053	hypothetical prote	347	6	1.9	249	2	J00133	hypothetical prote
275	6	1.9	198	2	D72603	hypothetical prote	348	6	1.9	250	2	A28564	hypothetical 26.4K
276	6	1.9	199	2	E69111	conserved hypochet	349	6	1.9	251	2	D71931	lymphocyte functio
277	6	1.9	200	1	Q0B827	BALF1 protein - hu	350	6	1.9	252	2	H75123	hypothetical prote
278	6	1.9	200	1	D64008	hypothetical prote	351	6	1.9	253	2	S75731	thiamin biosynthe
279	6	1.9	203	2	JC2207	Lam1 protein - tr	352	6	1.9	256	2	A84862	3-isopropylmalate
280	6	1.9	204	2	H83963	guanylate kinase g	353	6	1.9	257	2	JU0403	hypothetical 28.1K
281	6	1.9	205	1	A65094	YcH1 protein - Esc	354	6	1.9	257	2	A10844	glucitol operon re
282	6	1.9	205	2	F91121	hypothetical prote	355	6	1.9	257	2	S76930	hypothetical prote
283	6	1.9	205	2	E85966	hypothetical prote	356	6	1.9	257	2	A82680	conserved hypochet
284	6	1.9	205	2	B70168	conserved hypochet	357	6	1.9	258	2	C86286	protein Pg1.16 [I
285	6	1.9	205	2	D75178	hypothetical prote	358	6	1.9	258	2	S23106	p1A protein - shi
286	6	1.9	207	2	G75118	proteasome, chain	359	6	1.9	258	2	A81265	probable UDP-N-ace
287	6	1.9	207	2	A83540	phospholipase acce	360	6	1.9	259	2	G87083	probable rRNA meth
288	6	1.9	207	2	AG0055	probable membrane	361	6	1.9	259	2	E87537	hypothetical prote
289	6	1.9	209	2	H85644	hypothetical prote	362	6	1.9	259	2	G83057	hypothetical prote
290	6	1.9	209	2	C90785	probable anti-repr	363	6	1.9	260	2	G70619	probable 23s rnam
291	6	1.9	211	2	G82573	CDP-diacylglycerol	364	6	1.9	260	2	T26174	hypothetical prote
292	6	1.9	211	2	S21864	probable cysteine	365	6	1.9	260	2	A99290	hypothetical prote
293	6	1.9	214	2	G83692	hypothetical prote	366	6	1.9	261	2	B64746	YafT protein - Esc
294	6	1.9	215	2	S54293	keratin 6 - bovine	367	6	1.9	261	2	C97462	ATP synthase chain
295	6	1.9	215	2	C72739	hypothetical prote	368	6	1.9	262	2	S41753	acyl-lactyl-carrier
296	6	1.9	217	2	S10212	late 33K protease -	369	6	1.9	262	2	C97426	flagellar basal-bo
297	6	1.9	218	2	E82780	endonuclease III X	370	6	1.9	262	2	AC2644	flagellar basal-bo
298	6	1.9	218	1	AG3486	hemolysin III (imp	371	6	1.9	262	2	S60213	fomc protein - Str
299	6	1.9	219	1	HSX118	histone H1B - Afri	372	6	1.9	262	2	AD2819	conserved hypochet
300	6	1.9	219	2	A75628	response regulator	373	6	1.9	264	2	D83522	probable radical a
301	6	1.9	220	2	I51447	histone H1B - Afri	374	6	1.9	265	2	A43899	transcription fact
302	6	1.9	221	2	G90982	acetyltransferase	375	6	1.9	265	2	S77181	hypothetical prote
303	6	1.9	221	2	E85828	acetyl transferase	376	6	1.9	266	2	T35800	probable polyamine
304	6	1.9	222	2	E84283	TRK potassium upla	377	6	1.9	266	2	E69083	thiamin biosynthes
305	6	1.9	223	2	B97568	trp repressor bind	378	6	1.9	266	2	T07363	hypothetical prote
306	6	1.9	223	2	A12788	flavodoxin (import	379	6	1.9	266	2	D85506	hypothetical prote
307	6	1.9	223	2	C71287	probable hypochet	380	6	1.9	266	2	T37878	hypothetical prote
308	6	1.9	224	2	D72861	AcOrf-91 protein -	381	6	1.9	267	2	S08314	cell wall glycopro
309	6	1.9	224	2	T21788	hypothetical prote	382	6	1.9	267	2	F90028	hypothetical prote
310	6	1.9	226	2	T05976	hypothetical prote	383	6	1.9	268	2	T05819	hypothetical prote
311	6	1.9	226	2	T29210	hypothetical prote	384	6	1.9	268	2	A81424	cytolethal distend
312	6	1.9	226	2	B95362	protein (imported	385	6	1.9	269	2	A99396	conserved hypochet
313	6	1.9	228	2	G83115	30S ribosomal prot	386	6	1.9	270	2	T47853	hypothetical prote
314	6	1.9	228	2	G69405	conserved hypochet	387	6	1.9	272	2	T41070	conserved hypochet
315	6	1.9	229	2	A97349	uncharacterized co	388	6	1.9	273	2	AH0278	probable transp
316	6	1.9	229	2	F70568	hypothetical prote	389	6	1.9	273	2	AH2222	ABC nitrate transp
317	6	1.9	231	2	C86665	amino acid ABC tra	390	6	1.9	273	2	S64054	hypothetical prote
318	6	1.9	231	2	C70595	hypothetical prote	391	6	1.9	273	2	S67622	hypothetical prote
319	6	1.9	231	2	C64703	hypothetical prote	392	6	1.9	274	2	T47856	hypothetical prote
320	6	1.9	231	2	C71816	probable outer mem	393	6	1.9	275	2	T43004	hypothetical prote
321	6	1.9	232	2	A10019	phosphoglycolate p	394	6	1.9	276	2	S23891	secretory pathway

395	6	1.9	276	2	H75169	abc transporter PA
396	6	1.9	276	2	B87203	probable antiport
397	6	1.9	276	2	B75007	probable transacti
398	6	1.9	277	2	T21630	hypothetical prote
399	6	1.9	277	2	AB1883	nucleic transport
400	6	1.9	277	2	H87552	hypothetical prote
401	6	1.9	278	2	H75351	D-arabinitol 2-deh
402	6	1.9	278	2	A75427	hypothetical prote
403	6	1.9	279	2	P90080	hypothetical prote
404	6	1.9	280	2	T02471	hypothetical prote
405	6	1.9	280	2	E96643	hypothetical prote
406	6	1.9	281	2	D97100	uncharacterized pr
407	6	1.9	281	2	C84868	probable endochiti
408	6	1.9	281	2	A10461	sn-glycerol-3-phos
409	6	1.9	283	2	B87603	hypothetical prote
410	6	1.9	283	2	S13383	hydroxyproline-ric
411	6	1.9	284	2	S25140	serine proteinase
412	6	1.9	284	2	B41224	homeotic protein P
413	6	1.9	284	2	B90190	conserved hypochet
414	6	1.9	284	2	G86477	protein F1504.21 l
415	6	1.9	284	2	AE0314	conserved hypochet
416	6	1.9	286	2	T47064	hypothetical prote
417	6	1.9	286	2	A10230	probable binding-P
418	6	1.9	287	2	T48548	hypothetical prote
419	6	1.9	288	2	A13217	D-alanine amnotra
420	6	1.9	288	2	D84229	unknown L-arabinos
421	6	1.9	288	2	C83487	probable transcrip
422	6	1.9	288	2	S40173	Exon protein - Rhl
423	6	1.9	291	2	S48977	hypothetical prote
424	6	1.9	291	2	R82897	hypothetical prote
425	6	1.9	292	2	AC1460	sugar ABC transport
426	6	1.9	292	2	AD1097	sugar ABC transport
427	6	1.9	292	2	HE4149	conserved hypochet
428	6	1.9	293	2	S64649	beta-lactamase
429	6	1.9	293	2	B80885	hypothetical prote
430	6	1.9	293	2	G85753	hypothetical prote
431	6	1.9	293	2	B64880	probable multiple
432	6	1.9	293	2	F70724	hypothetical prote
433	6	1.9	293	2	B87632	carbonic anhydrase
434	6	1.9	293	2	G98163	cdp-P4a, A-D-gluc
435	6	1.9	293	2	AB3124	glycoyltransferas
436	6	1.9	294	2	C91105	hypothetical prote
437	6	1.9	294	2	F85950	hypothetical prote
438	6	1.9	294	2	G65078	transcription regu
439	6	1.9	294	2	A12016	hydroxymethylblian
440	6	1.9	295	2	I40810	hypothetical prote
441	6	1.9	295	2	B91166	hypothetical prote
442	6	1.9	295	2	B86012	hypothetical prote
443	6	1.9	295	2	S48102	xyloglucan endo-1,
444	6	1.9	295	2	AE0131	homeotic protein X
445	6	1.9	295	2	B97597	hypothetical prote
446	6	1.9	296	2	F75209	hypothetical prote
447	6	1.9	297	1	I39938	N-acetylmuramoyl-L
448	6	1.9	297	1	F65002	hypothetical prote
449	6	1.9	297	2	D91027	probable sugar nuc
450	6	1.9	297	2	B85871	Chemotaxis MotB pr
451	6	1.9	298	2	D87444	hypothetical prote
452	6	1.9	298	2	T25695	hypothetical prote
453	6	1.9	298	2	T22250	hypothetical prote
454	6	1.9	299	2	AC0507	transcription acti
455	6	1.9	299	2	AC0507	transcription acti
456	6	1.9	299	2	F84785	probable xylogluc
457	6	1.9	299	2	T33830	hypothetical prote
458	6	1.9	299	2	A70713	hypothetical prote
459	6	1.9	300	2	D83158	transcription acti
460	6	1.9	301	1	QOEC3R	transcription acti
461	6	1.9	301	1	B90631	transcription acti
462	6	1.9	302	1	B85482	transcription acti
463	6	1.9	302	1	F77021	proline dehydrogen
464	6	1.9	302	2	A13645	3-hydroxybutyrate
465	6	1.9	303	2	H83849	hypothetical prote
466	6	1.9	304	2	AG3324	UDP-N-acetylmurama
467	6	1.9	305	2	S20600	malm protein - Sal

malose operon per
probable endo alph
proline oxidase (p
arginase [imported
hypothetical prote
homeotic protein H
Br/SpM transposon
nodulin-26-like pr
hypothetical prote
hypothetical prote
cysteine synthase
Hpr (Ser/Thr) prot
hypothetical prote
phosphotriboylform
Hpr-P (Ser) kinase/
Hpr-P (Ser) kinase/
hypothetical prote
RNA-directed RNA p
probable flavoprot
FixB protein [impo
electron transfer
hypothetical prote
hypothetical prote
hypothetical prote
hypothetical prote
hypothetical prote
ADP-heptose syntha
sor-operon regulat
regulatory protein
hypothetical prote
protein B0238.11 l
cysteine synthase
GDP-D-mannose dehy
come-like protein
probable polysacch
catechol 2,3-dioxy
Caesyma F420-deps
NS,N10-methylamete
probable 60S ribos
UDP-N-acetylenolpy
hypothetical prote
pyruvate dehydroge
pyruvate dehydroge
pyruvate dehydroge
pyruvate dehydroge
phosphotransacetyl
phosphotransacetyl
hypothetical prote
6-phosphofructokin
acetoiln dehydrogen
acetoisomal enoyl
peroxisomal enoyl
conserved hypochet
hypothetical prote
hypothetical prote
phosphotransacetyl
hypothetical prote
chitinase (BC 3.2.
L-arabinose transp
phosphotransacetyl
nucleophyritogen d
probable extensin
homeotic protein H
hypothetical prote
lipoprotein [impor
flagellar motor sw
gal repressor - Ha
gene Tlx-1 protein

541	6	1.9	333	2	H89961	hypothetical prote	614	6	1.9	370	2	E82751	lipopolyaccharide
542	6	1.9	334	2	T43865	phosphate transpor	615	6	1.9	370	2	AB3334	daunorubicin resis
543	6	1.9	334	2	T22215	hypothetical prote	616	6	1.9	372	2	S76427	hypothetical prote
544	6	1.9	335	2	G82936	prolipoprotein dia	617	6	1.9	372	2	A98157	probable permease
545	6	1.9	335	2	T03229	probable chitinase	618	6	1.9	372	2	AH3130	ABC transporter, m
546	6	1.9	335	2	S03212	hypothetical prote	619	6	1.9	373	2	B82697	rod shape-determ
547	6	1.9	336	2	S70255	flagellin - Lyme d	620	6	1.9	373	2	A56392	beta-galactoside a
548	6	1.9	336	2	I40204	flagellin - Borrel	621	6	1.9	374	2	S21736	translocating chai
549	6	1.9	336	2	B72617	probable dipeptide	622	6	1.9	374	2	C91198	EGP protein [limp
550	6	1.9	336	2	B49683	prsh 3'-region pro	623	6	1.9	374	2	G86044	secreted protein E
551	6	1.9	336	2	A87300	TPR domain protein	624	6	1.9	374	2	C88734	protein F32B10.6 l
552	6	1.9	336	2	H75479	lipocate-protein 11	625	6	1.9	374	2	AC0983	hypothetical prote
553	6	1.9	336	2	T22767	hypothetical prote	626	6	1.9	375	2	A32827	fetuin precursor -
554	6	1.9	337	2	G95915	probable sugar ABC	627	6	1.9	375	2	A64183	xylose transport p
555	6	1.9	337	2	S37799	hypothetical prote	628	6	1.9	376	2	S04497	surface antigen PA
556	6	1.9	338	2	D90697	adhesin/invasin-1i	629	6	1.9	376	2	C90444	hypothetical prote
557	6	1.9	338	2	G85547	adhesin/invasin-1i	630	6	1.9	376	2	H82968	hypothetical prote
558	6	1.9	339	2	A50754	hypothetical prote	631	6	1.9	377	1	B0A058	virB10 protein - A
559	6	1.9	340	2	S40414	chitinase (BC 3.2.	632	6	1.9	377	2	AF3249	component of type
560	6	1.9	340	2	T45805	glucan endo-1,3-be	633	6	1.9	379	2	C75006	hypothetical prote
561	6	1.9	341	2	H90797	probable tellurium	634	6	1.9	381	2	S24611	latent membrane pr
562	6	1.9	342	2	S63356	probable membrane	635	6	1.9	381	2	S52985	cell wall protein
563	6	1.9	342	2	D49348	succinoglycan bios	636	6	1.9	381	2	F87553	aminoctransferase,
564	6	1.9	342	2	B64491	NS,M10-methylenete	637	6	1.9	382	2	T25538	hypothetical prote
565	6	1.9	342	2	G95975	glucosyltransferas	638	6	1.9	383	1	VGBEKG	glycoprotein precu
566	6	1.9	342	2	AD0346	probable transport	639	6	1.9	383	2	T35773	succinyl-Coa synth
567	6	1.9	342	2	AE2490	WP-repeat protein	640	6	1.9	384	2	S70989	rcsf protein - Myc
568	6	1.9	343	2	A56098	alpha(1,2)fucosylt	641	6	1.9	384	2	S74774	hypothetical prote
569	6	1.9	344	2	G64122	phosphoribosylform	642	6	1.9	385	2	E70154	DNA-directed DNA p
570	6	1.9	344	2	S45912	YAO2 protein - Yea	643	6	1.9	385	2	S68780	dopamine D1-like r
571	6	1.9	345	1	VABPA7	major capsid prote	644	6	1.9	385	2	C83174	hypothetical prote
572	6	1.9	345	2	E86202	hypothetical prote	645	6	1.9	385	2	A85607	hypothetical prote
573	6	1.9	346	2	F81351	probable D-alanine	646	6	1.9	386	1	LABBRJ	latent membrane pr
574	6	1.9	347	2	D96590	hypothetical prote	647	6	1.9	386	2	A82284	conserved hypochet
575	6	1.9	347	2	H75253	hypothetical prote	648	6	1.9	386	2	F83994	xylose operon tran
576	6	1.9	347	2	T50990	hypothetical prote	649	6	1.9	388	2	S25298	extensin (clone To
577	6	1.9	348	2	B75312	branched-chain ami	650	6	1.9	388	2	S77534	high-affinity bran
578	6	1.9	348	2	D88088	protein B0454.1 [i	651	6	1.9	389	2	H95003	aromatic amino aci
579	6	1.9	351	2	JU0056	flagellin - Serrat	652	6	1.9	389	2	C97876	apartate transami
580	6	1.9	351	2	T32717	hypothetical prote	653	6	1.9	389	2	D83139	hypothetical prote
581	6	1.9	354	2	T19856	hypothetical prote	654	6	1.9	390	1	Q0BE77	glycoprotein I pre
582	6	1.9	354	2	T24094	hypothetical prote	655	6	1.9	390	2	AG2549	hypothetical prote
583	6	1.9	354	2	B56392	beta-galactoside a	656	6	1.9	390	2	G88022	protein T27A1.4 [i
584	6	1.9	354	2	AG1324	glucosamine-fructo	657	6	1.9	390	2	E84066	hypothetical prote
585	6	1.9	354	2	A11695	weakly glucosamine	658	6	1.9	391	1	S02192	cellular tumor ant
586	6	1.9	355	2	AD0299	conserved hypochet	659	6	1.9	391	2	S55732	utroporphyrinogen d
587	6	1.9	355	2	A47039	nylon oligomer-deg	660	6	1.9	391	2	B83203	arginate o-acetyl
588	6	1.9	355	2	D95270	probable fatty aci	661	6	1.9	391	2	G83346	conserved hypochet
589	6	1.9	355	2	AE1642	membrane protein h	662	6	1.9	391	2	T20026	hypothetical prote
590	6	1.9	356	2	JH0289	class I histocampa	663	6	1.9	391	2	A13375	rare lipoprotein A
591	6	1.9	356	2	B89919	hypothetical prote	664	6	1.9	392	2	T44146	DR6 protein [limp
592	6	1.9	357	2	D84395	1-allo-threonine a	665	6	1.9	393	2	T01653	utroporphyrinogen d
593	6	1.9	357	2	S37925	hypothetical prote	666	6	1.9	393	2	T45434	probable succinyl-
594	6	1.9	359	2	S55032	dry-20 protein - C	667	6	1.9	393	2	S69537	hypothetical prote
595	6	1.9	359	2	B72290	branched chain ami	668	6	1.9	393	2	E97248	PIP-dependent amin
596	6	1.9	359	2	S42787	serine/threonine-r	669	6	1.9	393	2	B97447	hypothetical prote
597	6	1.9	362	2	JH0291	class I histocampa	670	6	1.9	394	2	B84830	probable uroporphy
598	6	1.9	362	2	A12878	class I histocampa	671	6	1.9	395	2	B75512	conserved hypochet
599	6	1.9	362	2	C97655	hypothetical prote	672	6	1.9	395	2	A54949	syndecan precursor
600	6	1.9	362	2	T33092	probable hemin ABC	673	6	1.9	395	2	S73732	MC306 homolog A05-
601	6	1.9	362	2	T33092	hypothetical prote	674	6	1.9	395	2	F72424	hypothetical prote
602	6	1.9	363	2	S76156	hypothetical prote	675	6	1.9	396	2	A91019	probable antibiotic
603	6	1.9	363	2	F84360	threonine synthase	676	6	1.9	396	2	C85863	probable antibioci
604	6	1.9	364	1	JE0292	threonine synthase	677	6	1.9	396	2	AF0791	probable transmem
605	6	1.9	365	2	T49733	probable homoserin	678	6	1.9	397	2	G95850	probable efflux pr
606	6	1.9	365	2	A43720	parathion hydrolas	679	6	1.9	397	2	C82992	hypothetical prote
607	6	1.9	365	2	B75398	hypothetical prote	680	6	1.9	397	2	T44477	hypothetical prote
608	6	1.9	365	2	A36047	galactoside 2-alph	681	6	1.9	397	2	T40838	probable transpor
609	6	1.9	366	1	W2MLR1	E2 protein - rhesu	682	6	1.9	398	1	VBBPA7	minor capsid prote
610	6	1.9	367	2	A13319	choleyglycine hyd	683	6	1.9	398	2	T20669	hypothetical prote
611	6	1.9	368	2	T24983	hypothetical prote	684	6	1.9	400	2	H84492	En/Spm transposon
612	6	1.9	368	2	G88636	protein W09G12.7 [685	6	1.9	402	2	JE0282	cell division prot
613	6	1.9	369	2	AC2665	conserved hypochet	686	6	1.9	402	2	AD3451	hypothetical prote

687	6	1.9	403	2	H86976	probable S-adenosyl	760	6	1.9	438	2	E55578	hypothetical prote
688	6	1.9	403	2	A71484	probable phosphogl	761	6	1.9	439	2	G75314	preproteic translo
689	6	1.9	404	1	A856CA	latent membrane pr	762	6	1.9	439	2	G70693	probable dnf prot
690	6	1.9	404	2	H72577	conserved hypocher	763	6	1.9	439	2	E82983	probable oxidoredu
691	6	1.9	405	2	H82010	probable periplasm	764	6	1.9	439	2	S51939	chitinase (EC 3.2.
692	6	1.9	405	2	G91020	probable racemase	765	6	1.9	440	2	G87444	8-amin-7-oxononan
693	6	1.9	405	2	F85864	probable racemase	766	6	1.9	440	2	A56693	receptor protein k
694	6	1.9	405	2	AB0793	probable MR-Wdr-1a	767	6	1.9	441	2	G64094	preproteic translo
695	6	1.9	405	2	E64995	hypothetical prote	768	6	1.9	441	2	T12011	cellulase (EC 3.2.
696	6	1.9	405	4	A61181	homeotic protein H	769	6	1.9	442	2	S50062	cell wall glycopro
697	6	1.9	407	2	S25841	beta-keetoacyl synt	770	6	1.9	443	2	D97618	probable mfs trans
698	6	1.9	407	2	T28795	hypothetical prote	771	6	1.9	443	2	AB2841	MFS permease/proti
699	6	1.9	407	2	H87341	Ompa family protei	772	6	1.9	443	2	DB1233	hypothetical prote
700	6	1.9	407	2	D95882	probable efflux pr	773	6	1.9	444	2	T84602	hydrogen peroxide-
701	6	1.9	408	2	A86652	lps biosynthesis p	774	6	1.9	444	2	AS5871	hypothetical prote
702	6	1.9	409	1	B0AC55	viabi10 protein - A	775	6	1.9	445	2	T30604	hypothetical prote
703	6	1.9	409	2	T23264	inositol 1,4,5-tri	776	6	1.9	445	2	T43319	hypothetical prote
704	6	1.9	410	2	S41648	hypothetical prote	777	6	1.9	447	1	T05070	hypothetical prote
705	6	1.9	411	2	S85919	translational elonga	778	6	1.9	447	2	S44133	hypothetical prote
706	6	1.9	411	2	F91074	probable flavodoxi	779	6	1.9	447	2	D85362	probable oxidoredu
707	6	1.9	411	2	F91074	probable flavodoxi	780	6	1.9	448	2	G81382	probable porin PA2
708	6	1.9	412	2	H95313	probable aromatic-	781	6	1.9	448	2	G83333	glycine receptor a
709	6	1.9	412	2	E84266	hypothetical prote	782	6	1.9	449	2	S12382	purine permease ho
710	6	1.9	414	2	C75461	hypothetical prote	783	6	1.9	449	2	E70016	gtp-binding protei
711	6	1.9	414	2	T33067	hypothetical prote	784	6	1.9	449	2	F82848	UDP-N-acetylglucos
712	6	1.9	414	2	E89467	protein ZK380.1 [1	785	6	1.9	450	2	AB2296	UDP-N-acetylglucos
713	6	1.9	415	2	G81864	probable membrane	786	6	1.9	451	2	C49898	glycine receptor a
714	6	1.9	415	2	S37689	paired box transcr	787	6	1.9	451	2	T41877	cellulobiose phospho
715	6	1.9	416	2	T45051	hypothetical prote	788	6	1.9	451	2	AF3453	ME53 orf139 - Bomb
716	6	1.9	416	2	H14433	hypothetical prote	789	6	1.9	451	2	T00981	probable glucosyltr
717	6	1.9	416	2	S11678	hypothetical prote	790	6	1.9	452	2	H88791	protein T13P2.1 [1
718	6	1.9	417	2	A70156	RNA polymerase sig	791	6	1.9	454	2	E75291	probable cell wall
719	6	1.9	417	2	H71662	UDP-N-acetylglucos	792	6	1.9	454	2	S66080	UDP-N-acetylglucos
720	6	1.9	418	2	S50652	hypothetical prote	793	6	1.9	457	2	A51669	streptogristin C (E
721	6	1.9	419	2	A12458	hypothetical prote	794	6	1.9	457	2	S20662	glycine receptor a
722	6	1.9	420	2	G97698	transcription term	795	6	1.9	457	2	H83662	DNA repair protein
723	6	1.9	420	2	AH2924	transcription term	796	6	1.9	458	1	S74543	transcription term
724	6	1.9	421	2	AF3252	sugar transport sy	797	6	1.9	458	2	T33777	glutathione-disulf
725	6	1.9	421	2	A56220	protein kinase (EC	798	6	1.9	458	2	B84948	glutathione reduct
726	6	1.9	421	2	G98158	hypothetical prote	799	6	1.9	459	2	AH2426	regulatory protein
727	6	1.9	421	2	AB3129	paired box transcr	800	6	1.9	460	2	G90035	tear protein (lmpo
728	6	1.9	422	1	A56674	probable efflux pr	801	6	1.9	460	2	C82964	glycosyltransferas
729	6	1.9	423	2	C81266	probable outer mem	802	6	1.9	462	2	AC2274	cytochrome P450 [1
730	6	1.9	423	2	H83484	protein W10G11.6 [803	6	1.9	462	2	GRECNK	nitrite extrusion
731	6	1.9	426	2	D88103	malp protein homol	804	6	1.9	463	1	H90844	nitrite extrusion
732	6	1.9	427	1	S63615	conserved hypocher	805	6	1.9	463	2	G85702	glutathione-disulf
733	6	1.9	427	2	S66046	site-specific DNA-	806	6	1.9	465	2	T39699	S-layer protein -
734	6	1.9	428	2	JH0634	conserved hypocher	807	6	1.9	465	2	A47023	glutathione-disulf
735	6	1.9	428	2	H69187	hypothetical prote	808	6	1.9	466	2	S39494	probable secreted
736	6	1.9	428	2	T32952	uncharacterized CB	809	6	1.9	466	2	A12231	nicotinicamide nicle
737	6	1.9	428	2	C97075	keatin 18, type I	810	6	1.9	467	2	S33181	inositol 1,4,5-tri
738	6	1.9	429	2	AB3550	hypothetical prote	811	6	1.9	468	2	E81876	replicative DNA he
739	6	1.9	430	2	S05481	hypothetical prote	812	6	1.9	468	2	T12725	terminase large ch
740	6	1.9	431	2	T29267	D-arabino 3,hexulo	813	6	1.9	468	2	E81145	hypothetical prote
741	6	1.9	431	2	E89357	probable aminopept	814	6	1.9	468	2	E69294	hypothetical prote
742	6	1.9	432	1	E64882	neutroal pentraxin	815	6	1.9	468	2	A12188	lysostaphin (EC 3.
743	6	1.9	432	2	T10894	hypothetical prote	816	6	1.9	470	2	G18452	variant surface gl
744	6	1.9	432	2	E70968	paired box transcr	817	6	1.9	470	2	E69461	4-hydroxybutyrate
745	6	1.9	432	2	S42234	protocorpyrinogen	818	6	1.9	472	2	AF3400	hypothetical coile
746	6	1.9	432	2	B70473	probable phosphotr	819	6	1.9	472	2	T39486	lipase - pseudomon
747	6	1.9	432	2	S56529	probable phosphotr	820	6	1.9	473	2	A43942	transcription fact
748	6	1.9	433	2	AF0667	Opasue-2 protein -	821	6	1.9	477	2	E46396	hydrolase (aux2) h
749	6	1.9	433	2	A34800	phosphatidylserine	822	6	1.9	477	2	164210	
750	6	1.9	437	2	T39592	hypothetical prote	823	6	1.9	477	2		
751	6	1.9	437	2	C72600	hypothetical prote	824	6	1.9	477	2		
752	6	1.9	438	2	E91240	probable membrane	825	6	1.9	477	2		
753	6	1.9	438	2	B86088	probable citrate p	826	6	1.9	477	2		
754	6	1.9	438	2	D70528	hypothetical prote	827	6	1.9	477	2		
755	6	1.9	438	2			828	6	1.9	477	2		
756	6	1.9	438	2			829	6	1.9	477	2		
757	6	1.9	438	2			830	6	1.9	477	2		
758	6	1.9	438	2			831	6	1.9	477	2		
759	6	1.9	438	2			832	6	1.9	477	2		

833	6	1.9	477	2	T46304	hypothetical prote
834	6	1.9	479	1	RDHUT	glutathione-disulf
835	6	1.9	479	1	B65051	hypothetical prote
836	6	1.9	479	2	E97337	aspartyl/asparagin
837	6	1.9	479	2	AD0845	probable flavoprot
838	6	1.9	479	2	A64117	serine-type D-Ala-
839	6	1.9	480	2	AC2440	hypothetical prote
840	6	1.9	480	2	H71430	hypothetical prote
841	6	1.9	480	2	JC7506	heparanase protein
842	6	1.9	480	2	JC7506	alpha-amylase (EC
843	6	1.9	484	2	JE0261	N-acetylglucosamin
844	6	1.9	484	2	B82110	conserved hypochet
845	6	1.9	485	2	T32538	hypothetical prote
846	6	1.9	485	2	T28076	hypothetical prote
847	6	1.9	486	2	S66097	cell-cycle protein
848	6	1.9	488	2	P84493	hypothetical prote
849	6	1.9	489	2	G64039	hypothetical prote
850	6	1.9	490	2	T41646	hypothetical prote
851	6	1.9	490	2	T39534	hypothetical prote
852	6	1.9	491	2	C46230	RNA-binding protei
853	6	1.9	491	2	S63306	LEFT protein homol
854	6	1.9	493	2	AD1398	drug-export protei
855	6	1.9	493	2	AG1773	drug-export protei
856	6	1.9	494	2	F72095	na-dependent trans
857	6	1.9	494	2	B86527	Na-dependent trans
858	6	1.9	496	2	D75261	conserved hypochet
859	6	1.9	496	2	T30976	hypothetical prote
860	6	1.9	498	1	VH1V6	nucleoprotein - in
861	6	1.9	498	1	VH1V3	nucleoprotein - in
862	6	1.9	498	1	A96541	probable hexokinase
863	6	1.9	499	1	S66677	thioredoxin-disulf
864	6	1.9	499	1	S26454	hypothetical prote
865	6	1.9	499	2	JC2062	transforming growt
866	6	1.9	500	1	S60829	probable aldehyde
867	6	1.9	500	2	A48053	phosphatidylserine
868	6	1.9	501	2	S48120	deoxyribodipyrimid
869	6	1.9	502	2	T35910	probable carboxyle
870	6	1.9	502	2	H82832	glucose-6-phosphat
871	6	1.9	502	2	T02746	cyclin A-like prot
872	6	1.9	502	2	A10151	probable bacteriop
873	6	1.9	503	2	A49432	activin receptor-1
874	6	1.9	503	2	JC2061	transforming growt
875	6	1.9	504	2	F70813	hypothetical prote
876	6	1.9	504	2	A81922	probable cell-surf
877	6	1.9	504	2	F72744	hypothetical prote
878	6	1.9	505	2	T28276	ORF MSV15 probabl
879	6	1.9	507	2	T31179	hypothetical prote
880	6	1.9	508	2	C87564	cell division proc
881	6	1.9	508	2	D81325	probable phosphate
882	6	1.9	509	2	A45992	activin A receptor
883	6	1.9	509	2	T59576	transforming growt
884	6	1.9	509	2	A46664	activin type I rec
885	6	1.9	509	2	T32201	hypothetical prote
886	6	1.9	510	1	S15620	l1 protein - human
887	6	1.9	510	1	S15627	l1 protein - human
888	6	1.9	510	1	T39601	serine carboxypept
889	6	1.9	510	2	A45338	connexin-56 - chic
890	6	1.9	511	2	S44275	dopamine receptor
891	6	1.9	511	2	JE0312	3-ketosteroid-deta
892	6	1.9	512	2	AD0416	probable type II s
893	6	1.9	514	2	A87266	senor histidine k
894	6	1.9	515	2	AH2996	glycerol-3-phospha
895	6	1.9	516	2	T50190	probable phosphati
896	6	1.9	518	1	P2W18	l2 protein - human
897	6	1.9	518	1	P2W15	l2 protein - human
898	6	1.9	518	1	P2W1B5	l2 protein - human
899	6	1.9	518	1	P2W1A7	l2 protein - human
900	6	1.9	518	1	S36542	l2 protein - human
901	6	1.9	519	2	S36471	l2 protein - human
902	6	1.9	519	2	A54590	GAGA transcription
903	6	1.9	520	2	S36489	l2 protein - human
904	6	1.9	520	2	S36495	l2 protein - human
905	6	1.9	521	2	B95899	probable ABC trans
906	6	1.9	521	2	S36571	l2 protein - human
907	6	1.9	521	2	A43319	chromaffin granule
908	6	1.9	521	2	T45608	hypothetical prote
909	6	1.9	523	1	VHNZMV	nucleocapsid prote
910	6	1.9	523	2	T30091	hypothetical prote
911	6	1.9	523	2	F71302	asparagine-tRNA l1
912	6	1.9	523	2	A38101	potassium channel
913	6	1.9	523	2	C84753	hypothetical prote
914	6	1.9	524	2	S46007	hypothetical prote
915	6	1.9	524	2	T06021	hypothetical prote
916	6	1.9	525	1	A48556	nucleocapsid prote
917	6	1.9	525	1	VHNZMH	nucleocapsid prote
918	6	1.9	525	1	VUN0272	nucleocapsid prote
919	6	1.9	525	2	A49601	nucleocapsid prote
920	6	1.9	525	2	T44824	hypothetical prote
921	6	1.9	525	2	AP3549	multidrug resistan
922	6	1.9	529	2	T35966	probable secreted
923	6	1.9	530	1	T44892	probable dihydrol
924	6	1.9	530	2	D82412	chaperonin, 60 Kd
925	6	1.9	530	2	T43089	transfer complex p
926	6	1.9	531	2	A98287	glpd gene homolog
927	6	1.9	531	2	A95918	probable oligopept
928	6	1.9	535	2	A46101	protein-tyrosine-p
929	6	1.9	535	2	F90418	ABC transporter, p
930	6	1.9	536	2	B46230	RNA-binding protei
931	6	1.9	538	2	AG3295	phosphoglycerate d
932	6	1.9	538	2	A57624	retinoblastoma pro
933	6	1.9	540	2	T47858	hypothetical prote
934	6	1.9	540	2	AC2353	peptide-chain-rele
935	6	1.9	541	2	T48811	hypothetical prote
936	6	1.9	541	2	A87595	amine oxidase, fla
937	6	1.9	544	2	S41626	spike protein chat
938	6	1.9	546	2	AC2368	hypothetical prote
939	6	1.9	546	2	T19139	hypothetical prote
940	6	1.9	546	2	T02029	DNA-binding protei
941	6	1.9	547	2	B66530	Similar to CCS1 [i
942	6	1.9	548	2	B46101	protein-tyrosine-p
943	6	1.9	548	2	T36881	probable transcrip
944	6	1.9	549	2	AH1019	Sodium,solute sym
945	6	1.9	549	2	S37814	MIF2 protein - yea
946	6	1.9	550	2	E95359	Probable ABC trans
947	6	1.9	550	2	B84484	Mutator-like trans
948	6	1.9	551	2	B82753	dihydrolipoamide S
949	6	1.9	551	2	T22121	hypothetical prote
950	6	1.9	551	2	T02139	calcium-dependent
951	6	1.9	553	1	T45596	telomere repeat-bi
952	6	1.9	554	2	JE0303	propenediol dehydr
953	6	1.9	554	2	A56111	glycerol dehydrata
954	6	1.9	554	2	AC0760	hypothetical prote
955	6	1.9	555	2	T26413	isocitrate lyase (
956	6	1.9	557	1	WZBY1	hypothetical prote
957	6	1.9	559	2	T22179	5'-nucleotidase (E
958	6	1.9	560	1	HMIY2	hemagglutinin prec
959	6	1.9	562	1	JX0153	hypothetical prote
960	6	1.9	562	2	AH2084	hypothetical prote
961	6	1.9	562	2	C71473	hypothetical prote
962	6	1.9	564	2	S36637	signal recognition
963	6	1.9	566	2	T45162	succinate dehydrog
964	6	1.9	568	1	HNN241	hemagglutinin-neur
965	6	1.9	568	2	D90525	hypothetical prote
966	6	1.9	568	2	B86364	hypothetical prote
967	6	1.9	570	2	T45712	Dpp receptor SAX p
968	6	1.9	570	2	D97738	hypothetical prote
969	6	1.9	573	2	AB2401	hypothetical prote
970	6	1.9	574	2	T00245	reverse transcript
971	6	1.9	577	2	T40297	membrane transport
972	6	1.9	578	2	T38775	hypothetical prote
973	6	1.9	579	2	A86681	amino acid permeas
974	6	1.9	579	2	G84583	CDC27/YWC2-like pr
975	6	1.9	580	2	AB0994	gamma-glutamyltran
976	6	1.9	580	2	D86426	hypothetical prote
977	6	1.9	581	1	ERADP5	fiber protein - hu
978	6	1.9	581	2	G83189	probable sodium/hy

```

979 6 1.9 581 2 A4551 insect-stage-speci
980 6 1.9 582 2 F71431 hypothetical prote
981 6 1.9 582 2 A50170 probable transport
982 6 1.9 583 2 F70592 probable lpgB prot
983 6 1.9 587 2 D95247 aspartyl-tRNA synt
984 6 1.9 587 2 A98112 aspartyl-tRNA 119
985 6 1.9 587 2 A96955 ferrous iron trans
986 6 1.9 589 2 AC1644 ABC transporter (A
987 6 1.9 591 2 S30145 ketol-acid reducto
988 6 1.9 591 2 T45681 ketol-acid reducto
989 6 1.9 592 2 S50972 RIB2 protein - yea
990 6 1.9 592 2 T51712 chreonine ammonia-
991 6 1.9 594 2 S36502 LI protein (altern
992 6 1.9 597 2 C83349 hypothetical prote
993 6 1.9 598 2 D71391 NADH2 dehydrogenas
994 6 1.9 598 2 F83508 flagellar M ring p
995 6 1.9 598 2 D84242 hypothetical prote
996 6 1.9 598 2 T20769 hypothetical prote
997 6 1.9 598 2 AB1236 intercalin protein
998 6 1.9 600 2 H81331 excludinase ABC c
999 6 1.9 601 2 T11451 NADH2 dehydrogenas
1000 6 1.9 601 2 A55921 serine/threonine k

```

ALIGNMENTS

RESULT 1

conserved hypothetical protein ykud - Bacillus subtilis

A69865
C/Species: Bacillus subtilis
C/Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 15-Oct-1999
C/Accession: A69865
R/Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berte
C.; Bron, S.; Bouillie, S.; Bruch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch
Nature 330, 249-256, 1997
A/Authors: Foulger, D.R.; Fritz, C.; Fujita, M.; Funo, S.; Galizzi, A.; Gall
Koeber, P.; Konigstein, G.; Hahne, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F
A/Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Manes
Y.M.; Ogawa, K.; Ogiwara, A.; Oudga, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portere
Nieder, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadele, Y.; Sato, T.; Scanlon
A/Authors: Schleich, S.; Schroefer, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Sero
T.; Winters, P.; Wipac, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A/Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Zumbstein, E.; Yoshida, K
A/Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A/Reference number: A69580; MUID:98044033; PMID:9384377
A/Accession: A69865
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-164 <KUN>
A/Cross-references: GB:Z9911; GB:AL009126; NID:G2633699; PIDN:CBM13277.1; PTD:6184994;
A/Experimental source: strain 168
C/Genetics:
A/Gene: ykud

Query Match 2.5%; Score 8; DB 2; Length 164;
Best Local Similarity 100.0%; Pred. No. 5.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 73 YOVQSGGT 80
Db 4 YOVQSGDT 11

RESULT 2

H84135
hypothetical protein BH388 [Imported] - Bacillus halodurans (strain C-125)
C/Species: Bacillus halodurans
C/Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C/Accession: H84135

R/Takami, H.; Nakase, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hiran
Nucleic Acids Res. 28, 4317-4331, 2000
A/Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and c
A/Reference number: A63650; MUID:20512582; PMID:11058132
A/Accession: H84135
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-227 <STO>
A/Cross-references: GB:AP001520; GB:BA000004; NID:G10176401; PIDN:BA07607.1; GSPDB:GN00
A/Experimental source: strain C-125
C/Genetics:
A/Gene: BH388

RESULT 3

S50385

hypothetical protein YLR301W - yeast (Saccharomyces cerevisiae)

N/Alternate names: hypothetical protein L8003.2

C/Species: Saccharomyces cerevisiae

C/Date: 13-Jan-1995 #sequence_revision 20-Feb-1995 #text_change 19-Apr-2002

C/Accession: S50385

R/Parley, A.

submitted to the EMBL Data Library, November 1994

A/Description: The sequence of S. cerevisiae cosmid 8003.

A/Reference number: S50385

A/Accession: S50385

A/Molecule type: DNA

A/Residues: 1244 <PAU>

A/Cross-references: EMBL:U17243; NID:G596030; PIDN:AAB67346.1; PTD:G596032; GSPDB:GN0001;

C/Genetics:

A/Gene: YLR301W

A/Cross-references: GSP:S0004292

A/Map position: 12R

C/Superfamily: Saccharomyces cerevisiae hypothetical protein YLR301W

Query Match 2.5%; Score 8; DB 2; Length 244;
Best Local Similarity 100.0%; Pred. No. 7.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 221 GSTVTSNG 228
Db 227 GSTVTSNG 234

RESULT 4

T52340

cell wall-plasma membrane linker protein homolog [Imported] - Arabidopsis thaliana

C/Species: Arabidopsis thaliana (mouse-ear cress)

C/Date: 24-Oct-2000 #sequence_revision 24-Oct-2000 #text_change 24-Oct-2000

C/Accession: T52340

R/Go, J.H.; Park, A.R.; Park, W.J.; Park, O.K.

Plant Mol. Biol. 41, 415-423, 1999

A/Title: Selection of Arabidopsis genes encoding secreted and plasma membrane proteins.

A/Reference number: Z55437; MUID:20064977; PMID:10598107

A/Accession: T52340

A/Status: preliminary; translated from GB/EMBL/DBD

A/Molecule type: mRNA

A/Residues: 1-306 <GOO>

A/Cross-references: EMBL:AF104328; PIDN:AAD11796.1

C/Genetics:

A/Note: CMLP

Query Match 2.5%; Score 8; DB 2; Length 306;
Best Local Similarity 100.0%; Pred. No. 9.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 168 PRTPPPV 175
 |||||
 Db 140 PRTPPPV 147

RESULT 5

751558
 Probable cell wall-plasma membrane linker protein PRP precursor - rape
 N:Alternate names: hybrid-proline-rich protein
 C:Species: Brassica napus (rape)
 C>Date: 27-Nov-1997 #sequence_revision 12-Dec-1997 #text_change 20-Jun-2000
 C:Accession: S71558
 R:Goodwin, W.; Pallat, J.A.; Jenkins, G.I.
 Plant Mol. Biol. 31, 771-781, 1996
 A>Title: Transcript of a gene encoding a putative cell wall-plasma membrane linker prot
 A:Reference number: S71558; MUID:96400032; PMID:8806408
 A:Accession: S71558
 A:Molecule type: DNA
 A:Residues: 1-376 <COO>
 A:Cross-references: EMBL:X94976; NID:g1155067; PIDN:CAA64425.1; PID:g1155068
 C:Superfamily: hydroxyproline-rich glycoprotein
 P:1-27/Domain: signal sequence #status predicted <SIG>
 F:28-376/Product: probable cell wall-plasma membrane linker protein PRP #status predicted

Query Match 2.5%; Score 8; DB 2; Length 376;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 168 PRTPPPV 175
 |||||
 Db 175 PRTPPPV 182

RESULT 6

S63211
 cytoskeleton assembly control protein SLA2 - yeast (Saccharomyces cerevisiae)
 N:Alternate names: MOP2 protein; protein N1102; protein YNL243w; UFG1 protein
 C:Species: Saccharomyces cerevisiae
 C>Date: 27-Apr-1996 #sequence_revision 03-May-1996 #text_change 29-Oct-1999
 C:Accession: S63211; S63209; S60673; S61932; S64652; S67355; S72073; S36354; S42278
 R:Saliz, J.E.; Baladron, V.; Del Rey, F.
 submitted to the Protein Sequence Database, April 1996
 A:Reference number: S63210
 A:Accession: S63211
 A:Molecule type: DNA
 A:Residues: 1-968 <SH1>

A:Cross-references: EMBL:Z71519; NID:g1302279; PIDN:CAA96149.1; PID:e239710; PID:g130228
 A:Experimental source: strain S288C
 R:Pandolfo, D.; De Antoni, A.; Lanfranchi, G.; Valle, G.
 submitted to the Protein Sequence Database, April 1996
 A:Reference number: S63188
 A:Accession: S63209
 A:Molecule type: DNA
 A:Residues: 820-968 <PAN>

A:Cross-references: EMBL:Z71519
 R:Holtzman, D.A.; Yang, S.; Drudin, D.G.
 J. Cell Biol. 122, 635-644, 1993
 A>Title: Synthetic-lethal interactions identify two novel genes, SLA1 and SLA2, that con
 A:Reference number: A40673; MUID:93328765; PMID:8335689
 A:Accession: B40673
 A>Status: nucleic acid sequence not shown

A:Molecule type: DNA
 A:Residues: 1-343, 'A', 345-968 <HO1>
 A:Cross-references: GB:Z22811; NID:g311412; PIDN:CAA80464.1; PID:g311413
 R:Yoon, H.; Donahue, T.F.
 submitted to the EMBL Data Library, March 1994
 A:Description: DNA sequence of ufg1 gene in yeast.

A:Reference number: S61932
 A:Accession: S61932
 A:Molecule type: DNA
 A:Residues: 327-343, 'A', 345-559, 'R', 561-886, 'NTVP', 891, 'MLPMHAEVWTTSW', 904 <YOO>

A:Cross-references: EMBL:U07938; NID:g501047; PIDN:AAA19161.1; PID:g501048
 A:Experimental source: strain 117-8A
 R:Na, S.; Hincapie, M.; McCusker, J.H.; Haber, J.E.
 J. Biol. Chem. 270, 6815-6823, 1995
 A>Title: MOP2 (SLA2) affects the abundance of the plasma membrane H(+) ATPase of Sacchar

A:Reference number: S64652; MUID:95204480; PMID:7896828
 A:Accession: S64652
 A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA
 A:Residues: 1-51, 'A', 53-343, 'A', 345-968 <NAs>
 A:Cross-references: EMBL:U12352; NID:g870731; PIDN:AAA74726.1; PID:g870732
 A>Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1995
 R:Pandolfo, D.; de Antoni, A.; Lanfranchi, G.; Valle, G.
 submitted to the EMBL Data Library, February 1996

A:Description: DNA sequence of cosmid 14-5 from chromosome XIV.
 A:Reference number: S67355
 A:Accession: S67355
 A:Molecule type: DNA
 A:Residues: 820-968 <PAN>
 A:Cross-references: EMBL:Z69381; NID:g1183970; PIDN:CAA93355.1; PID:e221951; PID:g1183971
 R:Pandolfo, D.; de Antoni, A.; Lanfranchi, G.; Valle, G.
 Yeast 12, 1071-1076, 1996

A>Title: The DNA sequence of cosmid 14-5 from chromosome XIV reveals 21 open reading fra
 A:Reference number: S72073; MUID:97051596; PMID:8896273
 A:Accession: S72073
 A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA
 A:Residues: 820-968 <PAF>
 A:Cross-references: EMBL:Z69381; NID:g1183970; PIDN:CAA93355.1; PID:e221951; PID:g1183971
 A>Note: the nucleotide sequence was submitted to the EMBL Data Library, February 1996
 C:Genetics:
 A:Gene: SGD:SLA2; UFG1; END4; MOP2
 A:Cross-references: MIPS:YNL243w; SGD:S0005187
 A:Map position: 14L
 C:Function:
 A:Description: essential for the assembly and function of the cortical cytoskeleton

C:Keywords: transmembrane protein
 F:771-787/Domain: transmembrane #status predicted <TM>

Query Match 2.5%; Score 8; DB 2; Length 968;
 Best Local Similarity 100.0%; Pred. No. 26;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 168 PRTPPPV 175
 |||||
 Db 295 PRTPPPV 302

RESULT 7

H86885
 50S ribosomal protein L24 [imported] - Lactococcus lactis subsp. lactis (strain IL1403)
 C:Species: Lactococcus lactis subsp. lactis
 C>Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
 C:Accession: H86885
 R:Boicoin, A.; Winkler, P.; Mauger, S.; Jallion, O.; Malarme, K.; Weissenbach, J.; Ehrlich
 Genome Res. 11, 731-753, 2001
 A>Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ser

A:Reference number: A86825; MUID:21255186; PMID:11337471
 A:Accession: H86885
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-101 <STO>
 A:Cross-references: GB:AE005176; PID:g12725141; PIDN:AAK06186.1; GSPDB:GN00146
 A:Experimental source: strain IL1403
 C:Genetics:
 A:Gene: rplX

C:Superfamily: Escherichia coli ribosomal protein L24

Query Match 2.2%; Score 7; DB 2; Length 101;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 279 VKTGDIV 285

|||||
Db 3 VKTGDIV 9

RESULT 8

S69849
hypothetical protein YMR290w-a - Yeast (Saccharomyces cerevisiae)
C/Species: Saccharomyces cerevisiae
C/Date: 23-Aug-1996 #sequence_revision 06-Sep-1996 #text_change 26-Aug-1999
C/Accession: S69849
R:Barrell, B.G.
submitted to the EMBL Data Library, August 1994
A/Reference number: S47445
A/Accession: S69849
A/Molecule type: DNA
A/Residues: 1-115 <BAR>
A/Cross-references: EMBL:X80836; MIPS:YMR290w-a
C/Genetics:
A/Map position: 13R
C/Superfamily: Saccharomyces hypothetical protein YMR290w-a

Query Match 2.2%; Score 7; DB 2; Length 115;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 298 PSGALF 304
Db 23 PSGALF 29

RESULT 9

T11639
ribosomal protein L17 - fission yeast (Schizosaccharomyces pombe)
N/Alternate names: 60S ribosomal protein
C/Species: Schizosaccharomyces pombe
C/Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 20-Jun-2000
R:Wood, V.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, December 1997
A/Reference number: Z17304
A/Accession: T11639
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-139 <WOO>
A/Cross-references: EMBL:AL021046
R:Lucas, M.; Galliard, C.; Lyne, M.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, January 1999
A/Reference number: Z21959
A/Accession: T40942
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-139 <LIC>
A/Cross-references: EMBL:AL035259; PIRN:CAA22864.1; GSPDB:GN00066; SPDB:SPCC1322.11
A/Experimental source: strain 972h-; cosmid c1322
C/Genetics: <RPL>
A/Map position: 1L
A/Accession: T27153
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-177 <WIL>
A/Cross-references: EMBL:AL032646; PIRN:CAA21686.1; GSPDB:GN00019; CESP:Y54E2A
A/Experimental source: clone Y54E2A

Query Match 2.2%; Score 7; DB 2; Length 139;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 241 ASNAGTV 247
Db 132 ASNAGTV 138

RESULT 10

T29781
hypothetical protein C50F2.7 - Caenorhabditis elegans
C/Species: Caenorhabditis elegans
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C/Accession: T29781
R:Du, Z.; Le, T.T.
submitted to the EMBL Data Library, November 1996
A/Description: The sequence of C. elegans cosmid C50F2.
A/Reference number: Z20684
A/Accession: T29781
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-141 <DUZ>
A/Cross-references: EMBL:U80445; PIRN:AA37800.1; GSPDB:GN00019; CESP:C50F2.7
A/Experimental source: strain Bristol N2; clone C50F2
C/Genetics:
A/Map position: 1
A/Introns: 19/3; 106/1

Query Match 2.2%; Score 7; DB 2; Length 141;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 222 STVTSNG 228
Db 58 STVTSNG 64

RESULT 11

T31531
hypothetical protein Y47D3A.20 - Caenorhabditis elegans
C/Species: Caenorhabditis elegans
C/Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Aug-2000
C/Accession: T31531
R:McCartney, L.
submitted to the EMBL Data Library, October 1999
A/Reference number: Z21043
A/Accession: T31531
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-161 <WIL>
A/Cross-references: EMBL:AL117202; PIRN:CA855080.1; CESP:Y47D3A.20
A/Experimental source: clone Y47D3A
C/Genetics:
A/Map position: 1
A/Introns: 42/1; 55/1; 83/2; 122/1
C/Superfamily: Caenorhabditis elegans hypothetical protein Y47D3A.20

Query Match 2.2%; Score 7; DB 2; Length 161;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 54 GGLAIGS 60
Db 108 GGLAIGS 114

RESULT 12

T27153
hypothetical protein Y54E2A.4 - Caenorhabditis elegans
C/Species: Caenorhabditis elegans
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 04-Mar-2000
C/Accession: T27153
R:Lloyd, C.
submitted to the EMBL Data Library, October 1998
A/Reference number: Z20319
A/Accession: T27153
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-177 <WIL>
A/Cross-references: EMBL:AL032646; PIRN:CAA21686.1; GSPDB:GN00019; CESP:Y54E2A
A/Experimental source: clone Y54E2A

C:Genetics:
 A:Gene: CESP:Y54E2A.4
 A:Map position: 1
 A:Insertions: 13/1; 86/3
 C:Superfamily: Caenorhabditis elegans hypothetical protein Y54E2A.4

Query Match 2.2%; Score 7; DB 2; Length 177;
 Best Local Similarity 100.0%; Pred. No. 57;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 STSGSGS 47
 |||||
 Db 62 STSGSGS 68

RESULT 13

probable membrane protein YP01754 [imported] - Yersinia pestis (strain CO92)

C:Species: Yersinia pestis
 C>Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Nov-2001
 C:Accession: A10213
 R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Tittball, R.W.; Holden, M.T.G.; Prentice, M.B.; deno-Terraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
 11, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrett,
 Nature 413, 523-527, 2001
 A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
 A:Reference number: AB0001; MUID:21470413; PMID:11586360

A:Accession: A10213
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-189 <KUR>
 A:Cross-references: GB:AL590842; PIDN:CAC90573.1; PID:g15979781; GSPDB:GN00175
 C:Genetics:
 A:Gene: YP01754
 C:Superfamily: Escherichia coli conserved hypothetical protein b1821

Query Match 2.2%; Score 7; DB 2; Length 189;
 Best Local Similarity 100.0%; Pred. No. 60;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 GLIRGVI 25
 |||||
 Db 40 GLIRGVI 46

RESULT 14

kinB sporulation signaling pathway activator kbaa - Bacillus subtilis

N:Alternate names: integral membrane protein kbaa
 C:Species: Bacillus subtilis
 C>Date: 31-Dec-1996 #sequence_revision 31-Dec-1996 #text_change 15-Oct-1999
 C:Accession: J06024; F69647
 R:Dartois, V.; Djavakhishvili, T.; Hoch, J.A.
 J. Bacteriol. 178, 1178-1186, 1996
 A:Title: Identification of a membrane protein involved in activation of the kinB pathway
 A:Reference number: J06024; MUID:96165276; PMID:8576055

A:Accession: J06024
 A:Molecule type: DNA
 A:Residues: 1-198 <DAR>
 A:Cross-references: GB:U23797; NID:g1143815; PIDN:AA44000.1; PID:g1143817
 R:Kunst, R.; Ogasawara, N.; Moser, I.; Albertini, A.M.; Alloni, G.; Azzevedo, V.; Berter
 C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capiano, V.; Carter, N.M.; Chd
 A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
 Nature 390, 249-256, 1997

A:Authors: Poulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Funa, S.; Galizzi, A.; Gallier
 lech, J.; Harwood, C.R.; Hentaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
 Koeter, P.; Konigslein, G.; Krogh, S.; Kumano, M.; Kurita, A.; Lapidus, A.; Lardinis,
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
 Y, M.; Ogawa, K.; Ogiwara, A.; Oudaga, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
 Rieger, M.; Rivolta, C.; Rocha, A.; Roche, B.; Rose, M.; Sadie, Y.; Sato, T.; Seanton,
 A:Authors: Schleich, S.; Schroeder, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Serot
 akenchi, M.; Tamakoshi, A.; Tanaka, T.; Terpetra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
 T.; Winters, P.; Wipac, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.

A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
 A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
 A:Reference number: A69580; MUID:98044033; PMID:9384377
 A:Accession: F69647

A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-198 <KUN>
 A:Cross-references: GB:Z99104; GB:AL009126; NID:g2632267; PIDN:CAB11932.1; PID:e1182089;
 A:Experimental source: strain 168
 C:Comment: This protein is an intrinsic membrane protein. It is highly hydrophobic and it
 C:Genetics:
 A:Gene: kbaa
 C:Keywords: membrane protein

Query Match 2.2%; Score 7; DB 2; Length 198;
 Best Local Similarity 100.0%; Pred. No. 62;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 LIRGVIT 26
 |||||
 Db 128 LIRGVIT 134

RESULT 15

protein T26F17.20 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)
 C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
 C:Accession: G86352
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
 ansen, N.F.; Hughes, B.; Huizar, L.
 Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.;
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maltli, R.; Marzilli,
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shim, P.; Southwick, A.M.; Sun, H.; Tallon, I
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A86141; MUID:21016719; PMID:11130712

A:Accession: G86352
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-208 <STO>
 A:Cross-references: GB:AE005172; NID:g6552738; PIDN:AAF16537.1; GSPDB:GN00141
 C:Genetics:
 A:Gene: T26F17.20
 A:Map position: 1

Query Match 2.2%; Score 7; DB 2; Length 208;
 Best Local Similarity 100.0%; Pred. No. 65;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 314 YVDEPLTV 320
 |||||
 Db 105 YVDEPLTV 111

RESULT 16

ded protein - Escherichia coli (strain K-12)

C:Species: Escherichia coli
 C>Date: 31-Dec-1998 #sequence_revision 17-Oct-1997 #text_change 01-Mar-2002
 C:Accession: H65003; E29803
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Berna, N.T.; Burland, V.; Riley, M.; Co
 A.; Rose, D.J.; Mau, B.; Siao, Y.
 Science 277, 1453-1462, 1997
 A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617; PMID:9278503
 A:Accession: H65003
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-211 <BLAT>

Jung, K.H.; Alam, M.; Freiltaa, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ehardt, H.; Lowe, T.M.; Li
A:Title: Genome sequence of *Haemophilus influenzae* species NRC-1.
A:Reference number: A84160; PMID:20504483; PMID:11016950
A:Accession: C84340
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-249 <STO>
A:Cross-references: GB:AE004437; NID:gl0581334; PIDN:AAZ0087.1; GSPDB:GN00138
C:Genetics:
A:Gene: trp2

Query Match
Best Local Similarity 100.0%; Pred. No. 76;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 215 GATVAVG 221
|||||
DB 60 GATVAVG 66

RESULT 22

C70151
hypothetical protein BB0412 - Lyme disease spirochete
C:Species: *Borrelia burgdorferi* (Lyme disease spirochete)
C:Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 08-Oct-1999
A:Accession: C70151
R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White
son, D.; Peterson, J.; Klevavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,
Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997
A:Authors: Smith, H.O.; Venter, J.C.
A:Title: Genomic sequence of a Lyme disease spirochete, *Borrelia burgdorferi*.
A:Reference number: A70100; PMID:9805943; PMID:9403685
A:Accession: C70151
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-259 <KLF>
A:Cross-references: GB:AE001146; GB:AE000783; NID:g2688312; PIDN:AAC66791.1; PID:g268832
A:Experimental source: strain B31

Query Match
Best Local Similarity 100.0%; Pred. No. 79;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 NNINSSY 105
|||||
DB 109 NNINSSY 115

RESULT 23

R89455
protein F55A4.4 [imported] - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Nov-2001
A:Accession: R89455
R:Anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode *C. elegans*: a platform for investigating biology
A:Reference number: A75000; PMID:99069613; PMID:9851916
A:Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C_ele
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A:Accession: F89455
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-269 <STO>
A:Cross-references: GB:chr_X; PIDN:AA07562.1; PID:gl519659; GSPDB:GN00028; CESP:F55A4.4
C:Genetics:
A:Gene: F55A4.4
A:Map position: X

Query Match
2.2%; Score 7; DB 2; Length 269;

Best Local Similarity 100.0%; Pred. No. 82;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 315 VDPLTVL 321
|||||
DB 14 VDPLTVL 20

RESULT 24

S57134
hypothetical protein YJR111c - yeast (*Saccharomyces cerevisiae*)
N:Alternate names: hypothetical protein J2009
C:Species: *Saccharomyces cerevisiae*
C:Date: 08-Jul-1995 #sequence_revision 08-Sep-1995 #text_change 19-Apr-2002
A:Accession: S57134
R:Rose, M.; Koetter, P.; Entian, K.D.
submitted to the Protein Sequence Database, September 1995
A:Reference number: S56848
A:Accession: S57134
A:Molecule type: DNA
A:Residues: 1-283 <ROS>
A:Cross-references: EMBL:Z49611; NID:gl015825; PIDN:CAA89641.1; PID:gl015826; GSPDB:GN00
C:Genetics:
A:Gene: MIPS:YJR111c
A:Cross-references: SGD:S0003872
A:Map position: 10R

Query Match
Best Local Similarity 100.0%; Pred. No. 85;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 NNINSSY 105
|||||
DB 191 NNINSSY 197

RESULT 25

H98073
hypothetical protein ABC-MSP [imported] - *Streptococcus pneumoniae* (strain R6)
C:Species: *Streptococcus pneumoniae*
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001
A:Accession: H98073
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; E
e, R.; Leblanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McHenry, S.; M
y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 185, 5703-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
A:Title: Genome of the Bacterium *Streptococcus pneumoniae* Strain R6.
A:Reference number: A97872; PMID:21429245; PMID:11544234
A:Accession: H98073
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-284 <KUR>
A:Cross-references: GB:AE007317; PIDN:AAL00421.1; PID:gl5459287; GSPDB:GN00174
C:Genetics:
A:Gene: ABC-MSP
C:Superfamily: maltose transport protein malG

Query Match
Best Local Similarity 100.0%; Pred. No. 86;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 254 MDGASIV 260
|||||
DB 181 MDGASIV 187

RESULT 26

D69184
UTP-glucose-1-phosphate uridylyltransferase (EC 2.7.7.9) MTH634 [similarity] - *Methanobac
C:Species: Methanobacterium thermoautotrophicum*
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 03-Jun-2002
A:Accession: D69184

R, Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.; Ki, S.; Church, G.M.; Danciger, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.; J. Bacteriol. 179, 7135-7155, 1997
 A:Title: Complete genome sequence of *Mechanobacterium thermoautotrophicum* Delta H: funct
 A:Accession: D69184
 A:Reference number: A69000; MUID:98037514; PMID:9371463
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-285 <MTH>
 A:Cross-references: GB:AE000844; GB:AE000666; NID:92621707; PIDN:AAB85140.1; PID:9262171
 A:Experimental source: strain Delta H
 A:Genetics: MTH634
 A:Gene: MTH634
 C:Superfamily: Escherichia coli UTP-glucose-1-phosphate uridylyltransferase
 C:Keywords: nucleotidyltransferase

Query Match
 Best Local Similarity 2.2%; Score 7; DB 2; Length 285;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 183 PPTVAP 189
 Db 189 PPTVAP 195

RESULT 27
 JH0204
 hypothetical 30.5K protein precursor - *Enterococcus faecalis* plasmid pAM-beta-1
 C:Species: *Enterococcus faecalis*
 C:Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 21-Jul-2000
 C:Accession: JH0204
 R:Swinfild, T.J.; Oultam, J.D.; Thompson, D.E.; Brem, J.K.; Minton, N.P.
 Gene 87, 79-90, 1990
 A:Title: Physical characterisation of the replication region of the *Streptococcus faecalis*
 A:Reference number: F10201; MUID:90236302; PMID:2110101
 A:Accession: JH0204
 A:Molecule type: DNA
 A:Residues: 1-288 <SWT>
 A:Cross-references: EMBL:X17092; NID:93023041; PIDN:AAC38600.1; PID:93023044
 C:Comment: A number of structural features of this protein suggest a localization at the
 C:Genetics:
 A:Genome: plasmid
 C:Superfamily: probable pheromone-responsive protein
 C:Keywords: plasmid replication
 F:1-24/Domain: signal sequence
 F:25-288/Product: hypothetical 30.5K protein #status predicted <STG>

Query Match
 Best Local Similarity 2.2%; Score 7; DB 2; Length 288;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 191 ATGSSGV 197
 Db 27 ATGSSGV 33

RESULT 28
 S28264
 hydroxyproline-rich glycoprotein - maize
 C:Species: *Zea mays* (maize)
 C:Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 24-Sep-1999
 C:Accession: S28264
 R:Raz, R.; Jose, M.; Moysa, A.; Martinez-Izquierdo, J.A.; Pulidomenech, P.
 Mol. Cell. Genet. 233, 252-259, 1992
 A:Title: Different mechanisms generating sequence variability are revealed in distinct
 A:Reference number: S22456; MUID:92293123; PMID:1603067
 A:Accession: S28264
 A:Molecule type: DNA
 A:Residues: 1-303 <RAZ>
 A:Cross-references: EMBL:X63134; NID:922332; PIDN:CAA44844.1; PID:922333
 C:Superfamily: hydroxyproline-rich glycoprotein
 C:Keywords: glycoprotein; hydroxyproline

Query Match
 Best Local Similarity 2.2%; Score 7; DB 2; Length 303;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 167 KPTPTTP 173
 Db 126 KPTPTTP 132

RESULT 29
 AE3325
 comt, competence lipoprotein [imported] - *Brucella melitensis* (strain 16M)
 C:Species: *Brucella melitensis*
 C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002
 C:Accession: AE3325
 R:Delvacchio, V.G.; Kapral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Ios, T.; Ivanova
 Proc Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
 A:Title: The genome sequence of the facultative intracellular pathogen *Brucella melite*
 A:Reference number: AD3252; PMID:11756688
 A:Accession: AE3325
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-309 <KUR>
 A:Cross-references: GB:AE008917; PIDN:AAL51768.1; PID:917982509; GSPDB:GN00190
 A:Experimental source: strain 16M
 C:Genetics: BME10587
 A:Map position: 1

Query Match
 Best Local Similarity 2.2%; Score 7; DB 2; Length 309;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 30 LAGCASK 36
 Db 47 LAGCASK 53

RESULT 30
 G84936
 kinase [imported] - *Buchnera* sp. (strain ABS)
 C:Species: *Buchnera* sp.
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 02-Mar-2001
 C:Accession: G84936
 R:Shigenobu, S.; Watanabe, H.; Hattori, M.; Sakaki, Y.; Ishikawa, H.
 Nature 407, 81-86, 2000
 A:Title: Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. A
 A:Reference number: A84930; MUID:20445173; PMID:10993077
 A:Accession: G84936
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-315 <STO>
 A:Cross-references: GB:AP000398; GSPDB:GN00144
 A:Experimental source: strain ABS
 C:Genetics: YB3052; BU060

Query Match
 Best Local Similarity 2.2%; Score 7; DB 2; Length 315;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 256 GASIVIQ 262
 Db 290 GASIVIQ 296

RESULT 31
 T26886
 hypothetical protein Y44A6.1 - *Caenorhabditis elegans*
 C:Species: *Caenorhabditis elegans*
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C/Accession: T26886
R/Accession: R.
submitted to the EMBL Data Library, January 1998
A/Reference number: Z20280
A/Accession: T26886
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-326 <MTL>
A/Cross-references: EMBL:AL021491; PIDN:CAA16373.1; GSPDB:GN00023; CESP:Y44A6B.1
A/Experimental source: clone Y44A6B
C/Genetics:
A/Gene: CESP:Y44A6B.1
A/Map position: 5
A/Introns: 33/1; 128/3; 170/3; 213/1; 265/3

Query Match 2.2%; Score 7; DB 2; Length 326;
Best Local Similarity 100.0%; Pred. No. 97;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 221 GSTVTSN 227
Db 211 GSTVTSN 217

RESULT 32
JQ0985
hydroxyproline-rich glycoprotein precursor - maize

C/Species: Zea mays (maize)
C/Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 24-Sep-1999
C/Accession: JQ0985; PS0215
R/Stefel, V.; Ruiz-Avila, L.; Raz, R.; Valles, M.F.; Gomez, J.; Pages, M.; Martinez-Izco
Plant Cell 2, 785-793, 1990
A/Title: Expression of a maize cell wall hydroxyproline-rich glycoprotein gene in early
A/Reference number: JQ0985; MUID:93005673; PMID:2152127
A/Accession: JQ0985
A/Molecule type: DNA
A/Residues: 1-328 <ST1>
A/Cross-references: GB:945164; NID:g257040; PIDN:AAB23539.1; PID:g257041
A/Accession: PS0215
A/Molecule type: protein
A/Residues: 27-43 <ST2>
A/Experimental source: leaf
C/Superfamily: hydroxyproline-rich glycoprotein
C/Keywords: glycoprotein
F/1-26/Domain: signal sequence #status predicted <SIG>
F/21-328/Product: hydroxyproline-rich glycoprotein #status predicted <MAT>

Query Match 2.2%; Score 7; DB 2; Length 328;
Best Local Similarity 100.0%; Pred. No. 97;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 167 KPPTTP 173
Db 121 KPPTTP 127

RESULT 33

T40981
probable quinone oxidoreductase - fission yeast (Schizosaccharomyces pombe)
C/Species: Schizosaccharomyces pombe
C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Jan-2000
C/Accession: T40981
R/Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Voicikert, G.
submitted to the EMBL Data Library, October 1998
A/Reference number: Z21961
A/Accession: T40981
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-329 <LYN>
A/Cross-references: EMBL:AL031966; PIDN:CAA21450.1; GSPDB:GN00066; SPDB:SPCC1442.16C
A/Experimental source: strain 972h-; cosmid c1442
C/Genetics:
A/Gene: SPCC285.01c; SPDB:SPCC1442.16c

A/Map position: 3
C/Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology

Query Match 2.2%; Score 7; DB 2; Length 329;
Best Local Similarity 100.0%; Pred. No. 97;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 279 VKTGDTV 285
Db 138 VKTGDTV 144

RESULT 34

140448
conserved hypothetical protein yyad (replication origin region) - Bacillus subtilis
C/Species: Bacillus subtilis
C/Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 21-Jul-2000
C/Accession: 140448; S66018; C70084; S18084
R/Ogasawara, N.; Yoshikawa, H.
Mol. Microbiol. 6, 629-634, 1992
A/Title: Genes and their organization in the replication origin region of the bacterial
A/Reference number: 140435; MUID:92204018; PMID:1552862
A/Accession: 140448
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-338 <RES>
A/Cross-references: EMBL:X62539; NID:g40020; PIDN:CAA44412.1; PID:g580907
R/Ogasawara, N.; Nakai, S.; Yoshikawa, H.
DNA Res. 1, 1-14, 1994
A/Title: Systematic sequencing of the 180 kilobase region of the Bacillus subtilis chrom
A/Reference number: S65967; MUID:96051385; PMID:7584024
A/Accession: S66018
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-338 <OGA>

A/Cross-references: EMBL:D26185; NID:g467326; PIDN:BA05224.1; PID:d1005766; PID:g467378
A/Note: the nucleotide sequence was submitted to the EMBL Data Library, December 1993
R/Kunst, F.; Ogasawara, N.; Moser, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertucci
C.; Bron, S.; Brouillet, S.; Bruchet, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A/Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Puma, S.; Gallizi, A.; Gallert
Iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.;
Koetter, P.; Koningsreid, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A.; Authors: lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Maunda, S.; Maueel,
Y.M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelie,
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon,
A/Authors: Schleich, S.; Schreier, R.; Scofield, F.; Sekiguchi, J.; Sekowska, A.; Seror,
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A/Authors: Yoshikawa, H.F.; Zumein, E.; Yoshikawa, H.; Danchin, A.
A/Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A/Reference number: A69580; MUID:98044033; PMID:9384377
A/Accession: C70084
A/Status: nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-338 <KUN>
A/Cross-references: GB:g29124; GB:AL009126; NID:g2636442; PIDN:CA16131.1; PID:g2636641
A/Experimental source: strain 168
C/Genetics:
A/Gene: yyad
A/Start codon: GNG

Query Match 2.2%; Score 7; DB 2; Length 338;
Best Local Similarity 100.0%; Pred. No. 16+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 19 GLIFGYI 25
Db 213 GLIFGYI 219

RESULT 35

T34158
 hypothetical protein C42D4.6 - *Caenorhabditis elegans*
 C/Species: *Caenorhabditis elegans*
 C/Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
 C/Accession: T34158
 R/Db: Z.; Le, T.
 Submitted to the EMBL Data Library, December 1995
 A/Description: The sequence of C. elegans cosmid C42D4.
 A/Reference number: Z21483
 A/Accession: T34158
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 1-339 <DUZ>
 A/Cross-references: EMBL:U41991; PIDN:AAA83342.1; CESP:C42D4.6
 C/Genetics:
 A/Introns: 135/1; 158/3

Query Match 2.2%; Score 7; DB 2; Length 339;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 237 DLIANSN 243
 Db 281 DLIANSN 287

RESULT 36

S19129
 proline-rich protein TRPP-F1 - tomato
 C/Species: *Lycopersicon esculentum* (tomato)
 C/Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 24-Sep-1999
 C/Accession: S19129; S16589;
 R/Db: Y.; Kenigsbuch, D.; Wachs, R.; Grussem, W.; Barg, R.
 Plant Mol. Biol. 18, 407-409, 1992
 A/Title: DNA sequence of the tomato fruit expressed proline-rich protein gene TRPP-F1
 A/Reference number: S19129; MUID:92119262; PMID:1731999
 A/Accession: S19129
 A/Molecule type: DNA
 A/Residues: 1-346 <SAL>
 A/Cross-references: EMBL:X61395; NID:G1669697; PIDN:CAA43666.1; PID:G19390
 A/Experimental source: cv. VPNT cherry
 R/Db: Y.; Wachs, R.; Grussem, W.; Barg, R.
 Plant Mol. Biol. 17, 149-150, 1991
 A/Title: Sequence coding for a novel proline-rich protein preferentially expressed in y
 A/Reference number: S16589; MUID:91329722; PMID:1868217
 A/Accession: S16589
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 34-346 <SAM>
 A/Cross-references: EMBL:X57076; NID:G19520; PIDN:CAA40361.1; PID:G19521
 C/Experimental source: cv. Arava; haplotype 2n
 C/Genetics:
 A/Accession: TRPP-F1
 A/Map position: 7
 C/Superfamily: hydroxyproline-rich glycoprotein

Query Match 2.2%; Score 7; DB 2; Length 346;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 169 TPTPPV 175
 Db 190 TPTPPV 196

RESULT 37

S22456
 hydroxyproline-rich glycoprotein - perennial teasine
 C/Species: *Zea diploperennis* (perennial teasine)
 C/Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 24-Sep-1999
 C/Accession: S22456; S18964
 R/Db: R.; Jose, M.; Moya, A.; Martinez-Izquierdo, J.A.; Puigdomenech, F.

Mol. Gen. Genet. 233, 252-259, 1992
 A/Title: Different mechanisms generating sequence variability are revealed in distinct re
 A/Reference number: S22456; MUID:92293123; PMID:1603067
 A/Accession: S22456
 A/Molecule type: DNA
 A/Residues: 1-350 <RAZ>
 A/Cross-references: EMBL:X64173; NID:G22091; PIDN:CAA45514.1; PID:G22092
 C/Superfamily: hydroxyproline-rich glycoprotein
 C/Keywords: glycoprotein; hydroxyproline

Query Match 2.2%; Score 7; DB 2; Length 350;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 167 KPTPTP 173
 Db 125 KPTPTP 131

RESULT 38

AC0302
 conserved hypothetical protein YPO2474 [imported] - *Yersinia pestis* (strain CO92)
 C/Species: *Yersinia pestis*
 C/Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001
 C/Accession: AC0302
 R/Db: Y.; Wren, B.W.; Thomson, N.R.; Tibball, R.W.; Holden, M.T.G.; Prentice, M.B.;
 deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, G.;
 11, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrett, E
 Nature 413, 523-527, 2001
 A/Title: Genome sequence of *Yersinia pestis*, the causative agent of plague.
 A/Reference number: AB0001; MUID:21470413; PMID:11586360
 A/Accession: AC0302
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-351 <KUR>
 A/Cross-references: GB:AL590842; PIDN:CAC91279.1; PID:G15980468; GSPDB:GNO0175
 C/Genetics:
 A/Accession: YPO2474

Query Match 2.2%; Score 7; DB 2; Length 351;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 16 KRGGLF 22
 Db 176 KRGGLF 182

RESULT 39

G71958
 hypothetical protein jhp0220 - *Helicobacter pylori* (strain J99)
 C/Species: *Helicobacter pylori*
 A/Variety: strain J99
 C/Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999
 C/Accession: G71958
 R/Db: A.; Ling, L.S.L.; Molt, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;
 Ives, C.; Gibson, R.; Metberg, D.; Mills, S.D.; Giang, Q.; Taylor, D.E.; Voyts, G.F.;
 Nature 397, 176-180, 1999
 A/Title: Genomic sequence comparison of two unrelated isolates of the human gastric path
 A/Reference number: A71800; MUID:99120557; PMID:9923682
 A/Accession: G71958
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-355 <ARN>
 A/Cross-references: GB:AE001460; GB:AE001439; NID:G4154733; PIDN:AA05807.1; PID:G4154746
 A/Experimental source: strain J99
 C/Genetics:
 A/Accession: jhp0220

Query Match 2.2%; Score 7; DB 2; Length 355;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 77 QGDTVSK 83
 |||||
 Db 287 QGDTVSK 293

RESULT 40

H83202
 poly(beta-d-mannuronate) lyase precursor AlgI PA3547 [imported] - Pseudomonas aeruginosa
 C/Species: Pseudomonas aeruginosa
 C/Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
 C/Accession: H83202
 R/Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
 .; Lory, S.; Olson, M.V.
 Nature 406, 959-964, 2000
 A/Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
 A/Reference number: A82950; MUID:2043737; PMID:10984043
 A/Accession: H83202
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-367 <STO>
 A/Cross-references: GB:AE004775; GB:AE004091; NID:G9949692; PIDN:AA06935.1; GSPDB:GN001
 A/Experimental source: strain PA01
 C/Genetics:
 A/Gene: algI; PA3547
 C/Superfamily: Pseudomonas aeruginosa poly(beta-D-mannuronate) lyase

Query Match 2.2%; Score 7; DB 2; Length 367;
 Best Local Similarity 100.0%; Freq. No. 1.1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 232 SGRDGL 238
 |||||
 Db 109 SGRDGL 115

Search completed: July 8, 2003, 11:21:37
 Job time : 65 secs

THIS PAGE BLANK (USPTO)


```

983 5 1.6 129 1 KV3M HUMAN P18136 homo sapien
984 5 1.6 129 1 KV4N MOUSE P01680 mus muscula
985 5 1.6 129 1 LYC3 ANANP P00796 anas platyr
986 5 1.6 129 1 LYC3 CHIMP P22320 chrysoloophu
987 5 1.6 129 1 LYC3 LOPE P22320 chrysoloophu
988 5 1.6 129 1 LYC3 PAVCR P19849 pavo cristat
989 5 1.6 129 1 LYC3 SYRRE P24533 symaticus
990 5 1.6 129 1 LYC3 SYRRO P81711 symaticus
991 5 1.6 129 1 MGSA THRTN O87B66 theromanaer
992 5 1.6 129 1 RMI4 ACACA P46767 acanthamoeb
993 5 1.6 129 1 RMI4 SCHPO O74416 schizosacch
994 5 1.6 129 1 Y207 AARPE O94192 aeropyrum p
995 5 1.6 130 1 BCCP STRMU P29337 streptococc
996 5 1.6 130 1 CLP1 DROME P02839 desophylla
997 5 1.6 130 1 YQJF ECOLI P42619 escherichia
998 5 1.6 131 1 AGSW VIDLV P79407 vulpes vulp
999 5 1.6 131 1 C79A LOCMI P45586 locusta mig
1000 5 1.6 131 1 C79B LOCMI P45587 locusta mig

```

ALIGNMENTS

RESULT 1
TS1_CANAL STANDARD; PRT; 196 AA.

AC 09750; (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Thiol-specific antioxidant protein.
GN TSA1.
OS Candida albicans (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5476;
[1]

SEQUENCE FROM N.A.
RA Marchais V., Cottin J.;
RT "Sequence of TSA-like gene in Candida albicans."
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: PHYSIOLOGICALLY IMPORTANT ANTIOXIDANT WHICH CONSTITUTES
CC AN ENZYMAIC DEFENSE AGAINST SULFUR-CONTAINING RADICALS. CAN
CC PROVIDE PROTECTION AGAINST A THIOL-CONTAINING OXIDATION SYSTEM BUT
CC NOT AGAINST AN OXIDATION SYSTEM WITHOUT THIOL.
CC -1- SUBUNIT: HOMODIMER, DISULFIDE-LINKED, UPON OXIDATION.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- PTM: THE CYS-48-SH GROUP IS THE PRIMARY SITE OF OXIDATION BY
CC H(2)O(2), AND THE OXIDIZED CYS-48 (PROBABLY CYS-SOH) RAPIDLY
CC REACTS WITH CYS-169-SH OF THE OTHER SUBUNIT TO FORM AN
CC INTERMOLECULAR DISULFIDE. THIS DISULFIDE IS SUBSEQUENTLY REDUCED
CC BY THIOREDOXIN

CC -1- SIMILARITY: BELONGS TO THE AHP/C-TSA FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: AF149421; AAD34017.1;
CC InterPro: IPR000866; AHP-C-TSA.
CC Pfam: PF00578; AHP-C-TSA; 1.
CC Antioxidant.
FT ACT_SITE 48 48 BY SIMILARITY.
FT ACT_SITE 169 169 BY SIMILARITY.
SQ SEQUENCE 196 AA; 21860 MM; 0BD7B6B4715DCDBB CRC64;

Query March 2.5%; Score 8; DB 1; Length 196;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 173 PVVQGPAP 180
DB 3 PVVQGPAP 10

RESULT 2

SIA2_YEAST STANDARD; PRT; 968 AA.

AC P33358; Q02434; (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 40, Last annotation update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE SIA2 protein (Transmembrane protein MOP2).
DE SIA2 OR ENDA OR MOP2 OR UFG1 OR YNL243W OR N1102.
GN Saccharomyces cerevisiae (Baker's yeast).
OS Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
[1]

SEQUENCE FROM N.A.
RA Holzman D.A., Yang S., Drubin D.G.;
RT "Synthetic-lethal interactions identify two novel genes, SIA1 and
RT SIA2, that control membrane cytoskeleton assembly in Saccharomyces
RT cerevisiae."
RL J. Cell Biol. 122:635-644 (1993).
[2]

SEQUENCE FROM N.A.
RA MEDLINE=95204480; PubMed=7896828;
RA Na S., Hincapié M., McCusker J.H., Haber J.E.;
RT "MOP2 (SIA2) affects the abundance of the plasma membrane H(+)-ATPase
RT of Saccharomyces cerevisiae."
RL J. Biol. Chem. 270:6815-6823 (1995).
[3]

SEQUENCE OF 327-968 FROM N.A.

RC STRAIN=117-8A;
RA Yoon H., Donahue T.F.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
[4]
SEQUENCE FROM N.A.
RA Saiz J.E., Baladron V., del Rey F.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
[5]

SEQUENCE OF 819-968 FROM N.A.
RA MEDLINE=97051596; PubMed=8896273;
RA Pandolfo D., de Antoni A., Ianfranchi G., Valle G.;
RT "The DNA sequence of cosmid 14-5 from chromosome XIV reveals 21 open
RT reading frames including a novel gene encoding a globin-like
RT domain."
RL Yeast 12:1071-1076 (1996).
[6]

-1- FUNCTION: REQUIRED FOR CELLULAR MORPHOGENESIS AND POLARIZATION OF
CC THE CORTICAL CYTOSKELETON. IT MIGHT ACT IN CONCERT WITH PROTEINS
CC SUCH AS CDG42 AND CDG43 TO LIMIT THE REGION OF CORTICAL PATCH
CC FORMATION TO THE CORTEX OF THE BUD. REQUIRED FOR THE ACCUMULATION
CC AND/OR MAINTENANCE OF PLASMA MEMBRANE H(+)-ATPASE ON THE CELL
CC SURFACE

CC -1- SIMILARITY: TO C.ELEGANS ZK370.3 AND TO THE C-TERMINUS OF TALIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: Z22811; CA80464.1;
CC EMBL: L12352; AAA74726.1;
CC EMBL: U07938; AAA19161.1;
CC EMBL: Z71519; CAA96149.1;
CC EMBL: Z71518; CAA96148.1;

```

DR EMBL; Z69381; CAA93355.1; -.
DR PIR; S36354; S36354.
DR PIR; B40673; B40673.
DR SGD; S0005187; SLA2.
DR InterPro; IPR001026; ENTH.
DR InterPro; IPR002558; ILMEQ.
DR Pfam; PF01417; ENTH; 1.
DR Pfam; PF01608; ILMEQ; 1.
DR ProDom; PD011820; ILMEQ; 1.
DR SMART; SM00273; ENTH; 1.
DR SMART; SM00307; ILMEQ; 1.
DR SMART; SM00307; ILMEQ; 1.
DR Structural protein; Cytoskeleton; Actin-binding; Transmembrane.
KM TRANSMEM 772 791
FT CONFLICT 52 52 P -> A (IN REF. 2).
FT CONFLICT 344 344 A -> R (IN REF. 4).
FT CONFLICT 560 560 A -> R (IN REF. 3).
FT CONFLICT 887 968 ECKSKDTDCRSISLGNHVKWGMIEDHSTSCQOQPLDTSTH
-> NTVPKMLPMHAEVWTVTSV (IN REF. 3).
SQ SEQUENCE 968 AA; 108910 MW; E592E09D8040CE9 CRC64;

Query Match 2.5%; Score 8; DB 1; Length 968;
Best Local Similarity 100.0%; Pred. No. 7.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 168 PRTPPVV 175
DB 295 PRTPPVV 302

RESULT 3
RL23 SCHPO STANDARD; PRT; 139 AA.
AC 042867;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 60S ribosomal protein L23.
(RPL23A OR RPL23 OR SPAC39.03) AND (RPL23B OR SPCC1322.11).
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
Schizosaccharomycetales; Schizosaccharomycetaceae;
Schizosaccharomyces.
NCBI_TaxID=4896;
[1]
RN NCBI_TaxID=4896;
RP SEQUENCE FROM N.A. (RPL23A AND RPL23B).
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
Sgouros J., Peat N., Hayles J., Baker S., Baaham D., Bowman S.,
Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
Gentles S., Goble A., Hamlin N., Harris P., Hildalgo J., Hodgson G.,
Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitch E.,
Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
Skellon J., Simmonds M., Squares R., Squares S., Stevens K.,
Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
Woodward J., Volckaert G., Aert R., Robben J., Gymnopoulos B.,
Welfens I., Vansireels E., Rieger M., Schaefer M., Mueller-Auer S.,
Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
Borzym K., Langer I., Beck A., Lehnach H., Reinhardt R., Pohl T.M.,
Eger P., Zimmermann W., Medler H., Wambutt R., Purnelle B.,
Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaire V., Motiller S.,
Gallbert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
Lucas M., Rochet M., Gallard C., Tallada V.A., Garzon A., Thode G.,
Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
Dominguez A., Revuelta J.L., Moreno S., Armstrong U., Forsburg S.L.,
Cerruti L., Lowe T., McCombie M.R., Paulsen I., Potashkin J.,
Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
"The genome sequence of Schizosaccharomyces pombe."

```

```

RL Nature 415:871-880(2002).
CC -1- MISCELLANEOUS: THERE ARE TWO GENES FOR L23 IN S. POMBE.
CC -1- SIMILARITY: BELONGS TO THE L1AP FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL021046; CAA15912.1; -.
DR EMBL; AL035259; CAA22864.1; -.
DR HSSP; P04450; 1WHT.
DR InterPro; IPR000218; Ribosomal_L14.
DR Pfam; PF00238; Ribosomal_L14; 1.
DR ProDom; PD001093; Ribosomal_L14; 1.
DR PROSITE; PS00049; RIBOSOMAL_L14; 1.
KM Ribosomal protein; Multigene family.
SQ SEQUENCE 139 AA; 14882 MW; D8C9A8903225FAD CRC64;

Query Match 2.2%; Score 7; DB 1; Length 139;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 241 ASNAQTV 247
DB 132 ASNAQTV 138

RESULT 4
YCF6_EUGER STANDARD; PRT; 161 AA.
AC YCF6_EUGER
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Hypothetical 18.6 kDa protein ycf67 (ORF161).
YCF67.
GN Euglena gracilis.
OS Euglena gracilis.
OC Chloroplast.
OC Eukaryota; Euglenozoa; Euglenida; Euglenales; Euglena.
NCBI_TaxID=3039;
[1]
RN NCBI_TaxID=3039;
RP SEQUENCE FROM N.A.
RC STRAIN=2;
RX MEDLINE=93347989; PubMed=8346031;
RA Hallick R.B., Hong L., Drager R.G., Favreau M.R., Monfort A.,
RA Great B., Spielmann A., Stutz E.;
RL "Complete sequence of Euglena gracilis chloroplast DNA.";
Nucleic Acids Res. 21:3537-3544(1993).
CC -1- SIMILARITY: BELONGS TO THE YCF67 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z11874; -; NOT ANNOTATED_CDS.
DR EMBL; X70810; CAA50117.1; -.
DR InterPro; IPR001700; RNA_pol_A_bac.
DR Pfam; PF01000; RNA_pol_A_bac; 1.
KM Chloroplast; Hypothetical protein.
SQ SEQUENCE 161 AA; 18606 MW; 377944A5C814949C CRC64;

Query Match 2.2%; Score 7; DB 1; Length 161;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 12 OKPIKRL 18
 DB 110 OKPIKRL 116

RESULT 5

KBAA_BACSU STANDARD; PRT; 198 AA.

ID KBAA_BACSU
 AC Plc449;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE KINB signaling pathway activation protein.

OS KBAA.
 OC Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

OX NCBI_TaxID=1423;

RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN=168 / JH642;
 RX MEDLINE=96165276; PubMed=8576055;
 RA Datois V., Djavakhishvili T., Hoch J.A.;
 RT "Identification of a membrane protein involved in activation of the
 KINB pathway to sporulation in Bacillus subtilis.";
 RL J. Bacteriol. 178:1178-1186(1996).

RN [2]
 RP SEQUENCE FROM N.A.

RC STRAIN=168;
 RX MEDLINE=97124188; PubMed=8969501;
 RA Yasumoto K., Liu H., Jeong S.M., Ohashi Y., Kakinuma S.,
 RA Tanaka K., Kawamura F., Yoshikawa H., Takahashi H.;
 RT "Sequence analysis of a 50 kb region between spo0H and rnh on the
 RT Bacillus subtilis chromosome.";
 RL Microbiology 142:3039-3046(1996).

RN [3]
 RP SEQUENCE FROM N.A.

RC STRAIN=168;
 RX MEDLINE=98044033; PubMed=9384377;
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
 RA Azavedo V., Berrtero M.G., Bessieres F., Bolochin A., Borcher S.,
 RA Borries R., Boudier L., Brans A., Braun M., Briggner S.C., Bron S.,
 RA Bruller S., Bruns C.V., Caldwell B., Capuano N.J., Carter N.M.,
 RA Choi S.K., Codan J.J., Conerton I.F., Cummings N.J., Daniel R.A.,
 RA Denicot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
 RA Enria K.D., Erttington J., Fabre C., Ferrari S., Foulger D.,
 RA Frits C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
 RA Gilm S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
 RA Guisepi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
 RA Hilbert H., Holasappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
 RA Kobayashi Y., Koester P., Koningsstein G., Krogh S., Kumano M.,
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
 RA Medina N., Mellado R.P., Mizuno M., Moesti D., Nakai S., Noback M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogata A., Oudega B., Park S.H.,
 RA Priesean E., Puig C., Purnelle D., Porwollik S., Prescott A.M.,
 RA Rieger M., Rivoita C., Rocha E., Roche B., Rose M., Sadate Y.,
 RA Sato T., Scallan E., Schleicher S., Schroeter R., Scoffone F.,
 RA Seiguchi J., Sekowaka A., Seror S.J., Seror P., Shin B.S., Soldo B.,
 RA Socorin A., Tacconi E., Takagi T., Takahashi H., Takemura K.,
 RA Takenuchi M., Tanakoshi A., Tanaka T., Tarpstra P., Tognoni A.,
 RA Tosiato V., Uchiyama S., Vanderbol M., Vannier F., Vassart A.,
 RA Viari A., Wambit R., Wedler E., Wedler H., Weitzenecker T.,
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumschein E., Yoshikawa H., Danchin A.,
 RT "The complete genome sequence of the Gram-positive bacterium Bacillus
 subtilis.";
 RL Nature 390:249-256(1997).

RN [4]
 RP SEQUENCE OF 1-91 FROM N.A.

RC STRAIN=168;
 RX MEDLINE=90257591; PubMed=2517635;

RA Yon J.R., Sammons R.L., Smith D.A.;
 RT "Cloning and sequencing of the gerD gene of Bacillus subtilis.";
 RL J. Gen. Microbiol. 135:3431-3445(1989).
 CC -1- FUNCTION: INVOLVED IN THE ACTIVATION OF THE KINB SIGNALING PATHWAY
 CC OF SPOULATION

CC -1- SUBCELLULAR LOCATION: Integral membrane protein.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; U23797; AAC44000.1; -;
 DR EMBL; D64126; BAA10936.1; -;
 DR EMBL; Z99104; CAB11932.1; -;
 DR EMBL; M27259; AAA22469.2; ALT INIT.

DR Subtilist; Bg10643; KbaA.
 KW Sporulation; Transmembrane; Complete proteome.

FT TRANSMEM 9 29 POTENTIAL.
 FT TRANSMEM 42 62 POTENTIAL.
 FT TRANSMEM 90 110 POTENTIAL.
 FT TRANSMEM 117 137 POTENTIAL.
 FT TRANSMEM 146 166 POTENTIAL.
 FT TRANSMEM 173 193 POTENTIAL.
 SQ SEQUENCE 198 AA; 22775 MW; 24917B51F38E0ACD CRC64;

Query Match 2.2%; Score 7; DB 1; Length 198;
 Best Local Similarity 100.0%; Pred. No. 19;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 20 LIRGVIT 26
 DB 128 LIRGVIT 134

RESULT 6

DEDD_ECOLI STANDARD; PRT; 220 AA.

ID DEDD_ECOLI
 AC P09549;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Dedd protein.
 GN DEDD OR B2314.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=562;

RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN=K12;
 RX MEDLINE=87308226; PubMed=3040734;
 RA Noret M.L., Marvel C.C., Tolan D.R.;
 RT "The hsr-purF region of the Escherichia coli K-12 chromosome."
 RT Identification of additional genes of the hsr and purF operons.";
 RL J. Biol. Chem. 262:12209-12217(1987).

RN [2]
 RP SEQUENCE FROM N.A.

RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blatter F.R., Plunkett G., Ili, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.D.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1234-1238(1997).

RN [3]
 RP SEQUENCE FROM N.A.

RC STRAIN=K12;
 RX MEDLINE=97349980; PubMed=9205837;

Yamamoto Y., Aiba H., Baba T., Hayashi K., Inada T., Isono K.,
 RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,
 RA Mizobuchi K., Mori H., Nakagawa S., Nakamura Y., Nishimoto H.,
 RA Oshima T., Oyama S., Satoh N., Sampei G., Satoh Y., Sivasubram S.,
 RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,
 RA Yamagata S., Horinouchi T.,
 RT "Construction of a contiguous 874-kb sequence of the *Escherichia coli*
 RT K12 genome corresponding to 50.0-68.8 min on the linkage map and
 RT analysis of its sequence features.",
 RL DNA Res. 4:91-113(1997).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M68934; AAA23967.1; ALT INIT.
 DR EMBL; AE000320; AAC75374.1; ALT INIT.
 DR EMBL; D90863; BAA16162.1; ALT INIT.
 DR EMBL; D90863; BAA16170.1; ALT INIT.
 DR PIR; E29803; XMECD.
 DR Ecogene; EG10218; dedb.
 KW Complete proteome.
 SQ SEQUENCE 220 AA; 22938 MW; 454529DCB8AC1644 CRC64;

Query Match 2.2%; Score 7; DB 1; Length 220;
 Best Local Similarity 100.0%; Pred. No. 21;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 178 PAPVAP 184
 DB 101 PAPVAP 107

RESULT 7

YPIVA_METTF STANDARD; PRT; 227 AA.
 AC P29577;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 01-APR-1993 (Rel. 25, Last annotation update)
 DE Hypochemical 24.7 kDa protein (ORF5A).
 OS Methanobacterium thermoformicicum.
 OG Plasmid pPVI.
 OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
 OC Methanobacteriaceae; Methanothermobacter.
 NCBI_TaxID=145262;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DSM 3848 / THF;
 RX MEDLINE=93126090; PubMed=1336177;
 RA Noelling J., van Rieden F.J.M., Eggen R.I.L., de Vos W.M.;
 RT "Molecular organization of related Archaeal plasmids encoding different
 RT restriction-modification systems in Methanobacterium
 RT thermoformicicum.",
 RL Nucleic Acids Res. 20:6501-6507(1992).
 CC -1- SIMILARITY: TO ORFS IN PEZ1.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X68366; CAA8429.1; --
 DR PIR; S26440; S26440.
 DR PIR; S30305; S30305.
 DR InterPro; IPR001434; DUF11.

DR Pfam; PF01345; DUF11.1.
 KW Hypochemical protein; Plasmid.
 SQ SEQUENCE 227 AA; 24681 MW; 2F3004C2519E849E CRC64;

Query Match 2.2%; Score 7; DB 1; Length 227;
 Best Local Similarity 100.0%; Pred. No. 21;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 36 KPTYNST 42
 DB 41 KPTYNST 47

RESULT 8

YJ81_YEAST STANDARD; PRT; 283 AA.
 ID YJ81_YEAST
 AC P47148;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Hypochemical 32.2 kDa protein in CPA2-NMF1 intergenic region.
 GN YJ81YC OR J2009.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Rose M., Koester P., Eutian K.D.;
 RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; Z49611; CAA89641.1; --
 DR SCD; S0003872; YJ81YC.
 KW Hypochemical protein.
 SQ SEQUENCE 283 AA; 32208 MW; CD49258DE05D384D CRC64;

Query Match 2.2%; Score 7; DB 1; Length 283;
 Best Local Similarity 100.0%; Pred. No. 26;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 NNINSSV 105
 DB 191 NNINSSV 197

RESULT 9

T212_LACLC STANDARD; PRT; 304 AA.
 ID T212_LACLC
 AC P50177;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Type II restriction enzyme L2ADCHI (EC 3.1.21.4) (Endonuclease
 DE L2ADCHI) (R.LADCHI) (L2ALI).
 GN L2ADCHI OR L2ALI.
 OS Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
 NCBI_TaxID=1359;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DCH-4;
 RX MEDLINE=95314272; PubMed=7793939;
 RA Moineau S., Walker S.A., Vedamuthu E.R., Vandenberg P.A.;
 RT "Cloning and sequencing of L2ADCHI restriction/modification genes

RT from Lactococcus lactis and relatedness of this system to the
RT Streptococcus pneumoniae DpnII system."
RL Appl. Environ. Microbiol. 61:2193-2202(1995).
CC -1- FUNCTION: RECOGNIZES THE DOUBLE-STRANDED UNMETHYLATED SEQUENCE
CC GATC AND CLEAVES BEFORE G-1
CC -1- CATALYTIC ACTIVITY: Endonucleolytic cleavage of DNA to give
CC specific double-stranded fragments with terminal 5'-phosphates.
CC -1- SIMILARITY: TO R.DNII AND R.MBOI.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; U16027; AAB06313.1; -.
DR REBASE; 2772; LIADCHI.
KM Hydrolyase; Endonuclease; Nuclease; Restriction system; Plasmid.
SQ SEQUENCE 304 AA; 34670 MW; C0C84547A28A20E4 CRC64;

Query Match 2.2%; Score 7; DB 1; Length 304;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 80 TVSKIAQ 86
Db 171 TVSKIAQ 177

RESULT 10
ID Y060_BUCAI STANDARD; PRT; 315 AA.
AC P57168;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical sugar kinase BU060.
GN BU060.
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
OS symbiotic bacterium).
OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
OX NCBI_TaxID=118099;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Tokyo 1998;
RX MEDLINE=20445173; PubMed=10993077;
RA Shigenobu S., Matanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT "Genome sequence of the endocellular bacterial symbiont of aphids
RT Buchnera sp. Aps.";
RL Nature 407:81-86(2000).
CC -1- SIMILARITY: BELONGS TO THE PKB FAMILY OF CARBOHYDRATE KINASES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; AP001118; BAB12783.1; -.
DR InterPro; IPR002173; PfkB.
DR Pfam; PF00294; PfkB; 1.
DR PROSITE; PS00583; PKB_KINASES_1; 1.
DR PROSITE; PS00584; PKB_KINASES_2; FALSE NEG.
KW Hypothetical protein; Transferase; Kinase; Complete proteome.
SQ SEQUENCE 315 AA; 35023 MW; 65B3707C601EFD25 CRC64;

Query Match 2.2%; Score 7; DB 1; Length 315;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 256 GASTVIO 262
Db 290 GASTVIO 296

RESULT 11
ID YVAD_BACSU STANDARD; PRT; 338 AA.
AC P37820;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein YVAD.
GN YVAD.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / CRK2000;
RX MEDLINE=92204018; PubMed=1552862;
RA Ogasawara N., Yoshikawa H.;
RT "Genes and their organization in the replication origin region of the
RT bacterial chromosome."
RL Mol. Microbiol. 6:629-634(1992).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; U16027; AAB06313.1; -.
DR REBASE; 2772; LIADCHI.
KM Hydrolyase; Endonuclease; Nuclease; Restriction system; Plasmid.
SQ SEQUENCE 304 AA; 34670 MW; C0C84547A28A20E4 CRC64;

Query Match 2.2%; Score 7; DB 1; Length 304;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 80 TVSKIAQ 86
Db 171 TVSKIAQ 177

RESULT 10
ID Y060_BUCAI STANDARD; PRT; 315 AA.
AC P57168;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical sugar kinase BU060.
GN BU060.
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
OS symbiotic bacterium).
OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
OX NCBI_TaxID=118099;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Tokyo 1998;
RX MEDLINE=20445173; PubMed=10993077;
RA Shigenobu S., Matanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT "Genome sequence of the endocellular bacterial symbiont of aphids
RT Buchnera sp. Aps.";
RL Nature 407:81-86(2000).
CC -1- SIMILARITY: BELONGS TO THE PKB FAMILY OF CARBOHYDRATE KINASES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; AP001118; BAB12783.1; -.
DR InterPro; IPR002173; PfkB.
DR Pfam; PF00294; PfkB; 1.
DR PROSITE; PS00583; PKB_KINASES_1; 1.
DR PROSITE; PS00584; PKB_KINASES_2; FALSE NEG.
KW Hypothetical protein; Transferase; Kinase; Complete proteome.
SQ SEQUENCE 315 AA; 35023 MW; 65B3707C601EFD25 CRC64;

Query Match 2.2%; Score 7; DB 1; Length 315;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X62539; CAA44412.1; -.
DR EMBL; D26185; BAA05224.1; -.
DR EMBL; Z99124; CAB16131.1; -.
DR PIR; S18084; S18084.
DR Subtilisin; BG10052; VYAD.
KM Hypothetical protein; Complete proteome.
SQ SEQUENCE 338 AA; 37721 MW; 8E5DE922917BBE84 CRC64;

Query Match 2.2%; Score 7; DB 1; Length 338;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 GLIFGVY 25
DB 213 GLIFGVY 219

RESULT 12
PRFL_LYCES STANDARD; PRT; 346 AA.
AC Q00451;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE 36.4 kDa proline-rich protein.
GN PRP-F1.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4081;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. VENT Cherry; TISSUE=Fruit;
RX MEDLINE=92119262; PubMed=1731999;
RA Salts Y., Wachs R., Kenigsbuch D., Gruissem W., Barg R.;
RT DNA sequence of the tomato fruit expressed proline-rich protein gene
RT PRP-F1 reveals an intron within the 3 untranslated transcript.;
RL Plant Mol. Biol. 18:407-409(1992).
RN [2]
RP SEQUENCE OF 34-346 FROM N.A.
RC STRAIN=cv. Arava;
RX MEDLINE=91329722; PubMed=1868217;
RA Salts Y., Wachs R., Gruissem W., Barg R.;
RT "Sequence coding for a novel proline-rich protein preferentially
RT expressed in young tomato fruit.";
RL Plant Mol. Biol. 17:149-150(1991).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X61395; CAA43666.1; -.
DR EMBL; X57076; CAA40361.1; -.
DR HSSP; P24337; 1HYP.
DR InterPro; IPR001768; Try/amy1_inhbt.
DR Pfam; PF00234; TRYD_alpha_amy1; 1.
SQ SEQUENCE 346 AA; 36375 MW; 604E58452F8F16E0 CRC64;

Query Match 2.2%; Score 7; DB 1; Length 346;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 169 TPTPPVV 175

```

```

DB 190 TPTPPVV 196

RESULT 13
ALGL_PSEAE STANDARD; PRT; 367 AA.
AC Q06749; Q57292;
DT 01-FEB-1995 (Rel. 31, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Alginatase lyase precursor (EC 4.2.2.3) (Poly(beta-D-mannuronate) lyase)
DE (Polymana) alginatase lyase.
GN ALGL OR PA3547.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93380656; PubMed=8370530;
RA Boyd A., Ghosh M., May T.B., Shinebarger D., Keogh R.,
RA Chakrabarty A.M.;
RT "Sequence of the algl gene of Pseudomonas aeruginosa and purification
RT of its alginatase lyase product.";
RL Gene 131:1-8(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=FRD1;
RX MEDLINE=93328683; PubMed=8335634;
RA Schiller N.L., Monday S.R., Boyd C.M., Keen N.T., Ohman D.E.;
RT "Characterization of the Pseudomonas aeruginosa alginatase lyase gene
RT (algl): cloning, sequencing, and expression in Escherichia coli.";
RL J. Bacteriol. 175:4780-4783(1993).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino E., Westbrock-Nadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
RN [4]
RP FUNCTION: Depolymerizes alginate by cleaving the beta-1,4
RN glycosidic bond. May enhance the production of alginate by
RN controlling the length of the polymer chain during export.
CC -1- CATALYTIC ACTIVITY: Eliminative cleavage of polysaccharides
CC containing beta-D-mannuronate residues to give oligosaccharides
CC with 4-deoxy-alpha-L-erythro-hex-4-enopyranuronosyl groups at
CC their ends.
CC -1- SUBCELLULAR LOCATION: Periplasmic.
CC -1- SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 5.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L14597; AAA71990.1; -.
DR EMBL; U27829; AAA91127.1; -.
DR EMBL; AE004775; AAG06935.1; -.
DR PIR; JN0777; JN0777.
DR Lyase; Signal; Periplasmic; Complete proteome.
KM STGNAL 1 27
FT CHAIN 28 367
FT CONFLICT 269 269 A -> P (IN REF. 1).

```

FT CONFLICT 337 341 KMLBA -> NACSRP (IN REF. 1).
SQ SEQUENCE 367 AA; 40829 MW; EAA3FE30032AB3BA CRC64;
Query Match 2.2%; Score 7; DB 1; Length 367;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 232 SGRDGL 238
Db 109 SGRDGL 115

RESULT 14
ID -RL3_YEAST STANDARD; PRT; 386 AA.
AC P14126; Q08459;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE 60S ribosomal protein L3 (YLI) (RPI) (Trichodermin resistance protein).
GN RPL3 OR TCM1 OR MAK8 OR YOR063W OR YOR29-14.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxId=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CLP1;
RX MEDLINE=83238226; PubMed=6305925;
RA "Nucleotide sequence of the tcm1 gene (ribosomal protein L3) of
RT Saccharomyces cerevisiae."
RL J. Bacteriol. 155:8-14(1983).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97278235; PubMed=9133743;
RA Valens M., Bohn C., Dagnan-Fornier B., Dang V., Bolotin-Fukuhara M.;
RT "The sequence of a 54.7 kb fragment of yeast chromosome XV reveals
RT the presence of two tRNAs and 24 new open reading frames."
RL Yeast 13:379-390(1997).
RN [3]
RP SEQUENCE OF 1-19.
RX MEDLINE=84038947; PubMed=6355773;
RA Oetka E., Higo K.I., Itoh T.;
RT "Yeast ribosomal proteins: VII. Cytoplasmic ribosomal proteins from
RT Schizosaccharomyces pombe."
RL Mol. Genet. 191:519-524(1983).
RN [4]
RP SEQUENCE OF 1-9.
RX MEDLINE=92184799; PubMed=1544921;
RA Takaura H., Tsunawasa S., Miyagi M., Warner J.R.;
RT "N2-terminal acetylation of ribosomal proteins of Saccharomyces
RT cerevisiae."
RL J. Biol. Chem. 267:5442-5445(1992).
CC -1- FUNCTION: THE L3 PROTEIN IS A COMPONENT OF THE LARGE SUBUNIT OF
CC CYTOPLASMIC RIBOSOMES.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- MISCELLANEOUS: A MUTANT ALLELE OF TCM1 CONFERES RESISTANCE TO
CC TRICHODERMIN, A TRICHOTECENE TOXIN PRODUCED BY PLANT-PATHOGENIC
CC FUNGI.
CC -1- SIMILARITY: BELONGS TO THE LBP FAMILY OF RIBOSOMAL PROTEINS.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; J01351; AAA88732.1; -;
DR EMBL; Z74971; CAA9256.1; -;

DR EMBL; Z70678; CAA94548.1; -;
DR PIR; S05843; RSBYAE.
DR SGD; S0005589; RPL3.
DR InterPro: IPR000597; Ribosomal_L3.
DR Pfam: PF00297; Ribosomal_L3; 1.
DR PROSITE: PS00474; RIBOSOMAL_L3; 1.
KW Ribosomal protein.
FT INIT MET 0
FT CONFLICT 254 254 W -> C (IN REF. 1).
SQ SEQUENCE 386 AA; 43626 MW; 8202CA08C071CB CRC64;
Query Match 2.2%; Score 7; DB 1; Length 386;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 159 TPPVVVV 165
Db 80 TPPVVVV 86

RESULT 15
ID -ALP_TRIHA STANDARD; PRT; 409 AA.
AC Q03420;
DT 01-OCT-1993 (Rel. 27, Created)
DT 15-OCT-1993 (Rel. 27, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Alkaline proteinase precursor (EC 3.4.21.-) (ALP).
GN PRP1.
OS Trichoderma harzianum.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreales; Microsporici Hypocreales; Trichoderma.
OX NCBI_TaxId=5544;
RN [1]
RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
RC STRAIN=IMI 206040;
RX MEDLINE=93316857; PubMed=8326868;
RA Geremia R.A., Goldman G.H., Jacobs D., Artiles W., Vila S.B.,
RA Van Montagu M., Herrera-Estrella A.;
RT "Molecular characterization of the proteinase-encoding gene, prp1,
RT related to mycoparasitism by Trichoderma harzianum."
RL Mol. Microbiol. 8:603-613(1992).
CC -1- FUNCTION: SERINE PROTEASE. SECRETED SPECIFICALLY DURING THE
CC MYCOPARASITIC PROCESS, WHICH IS INVOLVED IN THE DEGRADATION OF
CC PHYTOPATHOGEN CELL WALLS, MEMBRANES AND OF THE PROTEINS RELEASED
CC AFTER LYSIS OF THE HOST.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- INDUCTION: BY MYCEBIA, FUNGAL CELL WALLS, AND CHITIN, BUT ONLY
CC IN THE ABSENCE OF GLUCOSE.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M87518; AAA34211.1; -;
DR EMBL; M87516; AAA34209.1; -;
DR PIR; S32905; S32905.
DR HSSP; O99405; IMPT.
DR MEROPS; S08.00A.1;
DR InterPro: IPR000209; Peptidase_S8.
DR Pfam; PF00082; Peptidase_S8; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00136; SUBTILASE_ASP; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
KW Hydrolyase; Serine protease; Zymogen; Signal.
FT SIGNAL 1
FT PROPEP 21 120 POTENTIAL.

FT CHAIN 121 409 ALKALINE PHOSPHATASE.
 FT ACT SITE 161 161 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT SITE 192 192 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT SITE 353 353 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT CARBOHYD 252 252 N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE 409 AA, 42271 MW, D0F548EA2365C5B3 CRC64;

Query Match 2.2%; Score 7; DB 1; Length 409;
 Best Local Similarity 100.0%; Pred. No. 36;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 47 SHRTSGS 53
 Db 136 SHRTSGS 142

RESULT 16
 GAT3_BRARE STANDARD; PRT; 438 AA.
 ID GAT3_BRARE
 AC 091428;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Transcription factor GATA-3 (GATA Binding Factor-3).
 GN GATA3 OR GATA3.
 OS Brachydanio rerio (Zebrafish) (Danio rerio).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.
 OX NCBI_Taxid=7955;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96028520; Pubmed=7547465;
 RA Neave B., Rodaway A., Wilson S.W., Patient R., Holder N.;
 RT "Expression of zebrafish GATA 3 (gata3) during gastrulation and
 neurogenesis suggests a role in the specification of cell fate."; Mech. Dev. 51:169-182(1995).
 CC CC
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED DURING GASTRULATION IN THE VENTRAL
 CC REGION OF THE EMBRYO WHICH INCLUDES TISSUE PATED TO FORM THE NON-
 CC NEURAL ECTODERM.
 CC -1- SIMILARITY: CONTAINS 2 GATA-TYPE ZINC FINGERS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL, S80425; AAA93491.1; -;
 CC HSSP; P17679; 1GNF.
 CC ZFIN; ZDB-GENE-990415-82; gata3.
 CC InterPro; IPR000679; Znf_GATA.
 CC InterPro; IPR001164; hRIF_1like.
 CC Pfam; PF00320; GATA; 2.
 CC PRINTS; PR00619; GATAZNFINGER.
 CC SMART; SM00401; Znf_GATA; 2.
 CC PROSITE; PS00344; GATA_ZN_FINGER_1; 2.
 CC PROSITE; PS50114; GATA_ZN_FINGER_2; 2.
 CC Transcription regulation; Activator; DNA-binding; Zinc-finger;
 KM Nuclear protein.
 FT ZN_FING 256 280 GATA-TYPE 1.
 FT ZN_FING 310 334 GATA-TYPE 2.
 FT DOMAIN 69 72 POLY-PRO.
 FT DOMAIN 130 135 POLY-SER.
 FT DOMAIN 235 238 POLY-SER.
 FT DOMAIN 315 319 POLY-THR.
 SQ SEQUENCE 438 AA, 47590 MW, 5286C29C9775653C CRC64;

Query Match 2.2%; Score 7; DB 1; Length 438;
 Best Local Similarity 100.0%; Pred. No. 38;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 41 STSGSGS 47
 Db 161 STSGSGS 167

RESULT 17
 GATA_MYCPN STANDARD; PRT; 478 AA.
 ID GATA_MYCPN
 AC P75534;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Glutamyl-tRNA(Gln) amidotransferase subunit A (EC 6.3.5.-) (Glu-ADT
 DE subunit A).
 GN GATA OR MEN237 OR MP594.
 OS Mycoplasma pneumoniae.
 OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
 OX NCBI_Taxid=2104;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 29342 / M129;
 RX MEDLINE=97105885; Pubmed=8948633;
 RA Himmelfreid R., Hilbert H., Plagens H., Pirk E., Li B.-C.,
 RA Hermann R.;
 RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
 RT pneumoniae."; Res. 24:4420-4449(1996).
 RL Nucleic Acids.
 CC CC
 CC -1- FUNCTION: FURNISHES A MEANS FOR FORMATION OF CORRECTLY CHARGED
 CC GLN-tRNA(GLN) THROUGH THE TRANSAMINATION OF MISACYLATED GLU-
 CC tRNA(GLN) IN ORGANISMS WHICH LACK GLUTAMINYL-tRNA SYNTHETASE. THE
 CC REACTION TAKES PLACE IN THE PRESENCE OF GLUTAMINE AND ATP THROUGH
 CC AN ACTIVATED GAMMA-PHOSPHO-GLU-tRNA(GLN) (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: ATP + L-glutamyl-tRNA(Gln) + L-glutamine = ADP
 CC + phosphate + L-glutamyl-tRNA(Gln) + L-glutamate.
 CC -1- SUBUNIT: HETEROTRIMER OF A, B AND C SUBUNITS (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE AMIDASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AE000058; AAB96242.1; -;
 CC InterPro; IPR000120; Amidase.
 CC InterPro; IPR004412; Gata.
 CC Pfam; PF01425; Amidase; 1.
 CC TIGRFAMs; TIGR00132; gata; 1.
 CC PROSITE; PS00571; AMIDASSS; 1.
 CC Protein biosynthesis; Ligase; Complete proteome.
 SQ SEQUENCE 478 AA, 53229 MW, F228E9A84F641071 CRC64;

Query Match 2.2%; Score 7; DB 1; Length 478;
 Best Local Similarity 100.0%; Pred. No. 41;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 308 ISRNGVY 314
 Db 187 ISRNGVY 193

RESULT 18
 KIR2_HUMAN STANDARD; PRT; 505 AA.
 ID KIR2_HUMAN
 AC P36856; O15479; O15480; O15481; O15482;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Serine/threonine-protein kinase receptor R2 precursor (EC 2.7.1.37)

DE (SKR2) (Activin receptor-like kinase 4) (ALK-4) (ACTR-1B).
 DR ACVR1B OR ACVR1L4.
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 ON NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=9330967; PubMed=8397373.
 RA ten Dijke P., Ichijo H., Franzen P., Schulz P., Sarae J.,
 RA Toyoshima H., Heldin C.-H., Miyazono K.;
 RT Activin receptor-like kinases: a novel subclass of cell-surface
 RT receptors with predicted serine/threonine kinase activity.";
 RL Oncogene 8:2879-2887(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Kidney;
 RX MEDLINE=94254839; PubMed=8196624;
 RA Carcamo J., Wells F.M., Ventura F., Wieser R., Wiana J.L.,
 RA Attiano L., Massague J.;
 RT "Type I receptors specify growth-inhibitory and transcriptional
 RT responses to transforming growth factor beta and activin.";
 RL Mol. Cell. Biol. 14:3810-3821(1994).
 RN [3]
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
 RC TISSUE=Liver;
 RX MEDLINE=94336666; PubMed=8058741;
 RA Xu J., Matsuzaki K., McKeenan K., Wang F., Kan M., McKeenan W.L.;
 RT "Genomic structure and cloned cDNAs predict that four variants in the
 RT kinase domain of serine/threonine kinase receptors arise by
 RT Proc. Natl. Acad. Sci. U.S.A. 91:7957-7961(1994).
 RN [4]
 RP SEQUENCE FROM N.A. (ISOFORM 1/SKR2-1).
 RC TISSUE=Eye;
 RX Strauberg R.;
 RT Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS; 1/SKR2-1 (SHOWN HERE), 2/SKR2-2
 CC AND 3/SKR2-3; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN MANY TISSUES, MOST STRONGLY IN
 CC HUMAN KIDNEY, PANCREAS, BRAIN, LUNG, AND LIVER.
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 CC TGF-RECEPTOR SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>
 CC or send an email to license@ebi.ac.uk).
 CC -----
 DR EMBL: Z22536; CAAB0258.1; -
 DR EMBL: U14722; AAA50246.1; -
 DR EMBL: L10125; AAA50555.1; -
 DR EMBL: L10126; AAA50556.1; -
 DR EMBL: L31846; AAA53349.1; -
 DR EMBL: L31848; AAA53350.1; -
 DR EMBL: L31848; AAA53351.1; -
 DR EMBL: BC000254; AAH00254.1; -
 DR EMBL: PIR; S37184; S37184.
 DR HSSP; P36897; ITBI.
 DR Genew; HGNC:172; ACVR1B.
 DR MIM; 601300; -
 DR InterPro; IPR000472; Activin_rec.
 DR InterPro; IPR000719; Euk_pkinase.
 DR InterPro; IPR002290; Ser_thr_kinase.
 DR InterPro; IPR003605; TGFbeta_GS.
 DR Pfam; PF00069; Pkinase; 1.
 DR Pfam; PF01064; Activin_rec; 1.

DR Prodom; PD0000001; Euk_pkinase; 1.
 DR SMART; SM00467; GS; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
 DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
 KW Receptor; Transferase; Serine/threonine-protein kinase; ATP-binding;
 KW Transmembrane; Glycoprotein; Signal; Alternative splicing;
 KW Polymorphism.
 FT CHAIN 1 23
 FT SIGNAL 1 23
 FT DOMAIN 24 126
 FT TRANSMEM 127 149
 FT DOMAIN 150 505
 FT DOMAIN 207 497
 FT NP_BIND 213 221
 FT BINDING 234 234
 FT ACT_SITE 335 335
 FT CARBOHYD 43 43
 FT VARSPPLIC 422 505
 FT FT
 FT VARSPPLIC 465 505
 FT FT
 FT VARIANT 408 408
 FT FT
 FT CONFLICT 56 56
 FT CONFLICT 222 223
 FT SEQUENCE 505 AA; 56806 MW; 40AC65CAAC7573 CRC64;
 SQ
 Query Match 2.2%; Score 7; DB 1; Length 505;
 Best Local Similarity 100.0%; Pred. No. 43;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 41 STSGSGS 47
 DB 187 STSGSGS 193
 AC P80202;
 ID KIR2_RAT STANDARD; PRT; 505 AA.
 AC P80202;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Serine/threonine-protein kinase receptor R2 precursor (EC 2.7.1.37)
 DE (SKR2).
 GN ACVR1B OR ACVR1L4.
 OS Rattus norvegicus (Rat).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Muridae; Murinae; Rattus.
 ON NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Sprague-Dawley; TISSUE=urogenital ridge;
 RX MEDLINE=9337378; PubMed=8395914;
 RA He W.-W., Gustafson M., Hirobe S., Donahoe P.;
 RT "Developmental expression of four novel serine/threonine kinase
 RT receptor homologues to the activin/transforming growth factor-beta
 RT type II receptor family.";
 RL Dev. Dyn. 196:133-142(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placental;
 RX MEDLINE=95113007; PubMed=7813622;
 RA Takumi T., Moustakas A., Lin H.Y., Lodish H.F.;
 RT "Molecular characterization of a type I serine-threonine kinase
 RT receptor for TGF-beta and activin in the rat pituitary tumor cell
 RT line GH3.";
 RL Exp. Cell Res. 216:208-214(1995).

```

CC -1- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: UROGENITAL RIDGE, TESTIS, OVARY, BRAIN AND
CC LUNGS.
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC TGRB RECEPTOR SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; S76466; AAB3045.1; -.
DR HSSP; P36897; 1TBI.
DR InterPro; IPR000472; Activin_rec.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002290; Ser_thr_kinase.
DR InterPro; IPR003605; TGFbeta_GS.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF01064; Activin_recp; 1.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00467; GS; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00108; PROTEIN KINASE ST; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR Receptor; Transferrase; Serine/threonine-protein kinase; ATP-binding;
KW Transmembrane; Glycoprotein; Signal.
FT CHAIN 1 23
FT SIGNAL 1 23
FT SERINE/THREONINE-PROTEIN KINASE RECEPTOR
FT R2.
FT EXTRACELLULAR (POTENTIAL).
FT DOMAIN 24 126
FT TRANSMEM 127 149
FT DOMAIN 207 497
FT NP_BIND 213 221
FT BINDING 234 234
FT ACT_SITE 335 335
FT CARBOHYD 43 43
FT SEQUENCE 505 AA; 56804 MW; 377F4A5C867B3860 CRC64;
SQ
Query Match 2.2%; Score 7; DB 1; Length 505;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 41 STSGSGS 47
DB 187 STSGSGS 193

```

```

RN [2]
RX SEQUENCE FROM N.A.
RX MEDLINE=92090725; Pubmed=1752438;
RA Macdonald P.M., Luk S.K.-S., Kilpatrick M.;
RT "Protein encoded by the exuperantia gene is concentrated at sites of
RT bicoid mRNA accumulation in Drosophila nurse cells but not in oocytes
RT or embryos."
RT Genes Dev. 5:2455-2466(1991).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RX MEDLINE=20196006; Pubmed=10731132;
RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortan J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballaw R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolhakov S.,
RA Borokova D., Borchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Fiesel C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostali M., Houston K.A., Howland T.J., Wei M.-H., Idegawa C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Mostreil A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Snie B.C., Siden-Kiamos I., Simpson W., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirekas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RT Science 287:2185-2195(2000).
CC -1- FUNCTION: ENSURES THE PROPER LOCALIZATION OF THE MRNA OF THE
CC BICOID GENE TO THE ANTERIOR REGIONS OF THE OOCYTE THUS PLAYING
CC A FUNDAMENTAL ROLE IN THE ESTABLISHMENT OF THE POLARITY OF THE
CC OOCYTE. MAY BIND THE BCD MRNA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; S72757; AAB20673.1; -.
DR EMBL; S72363; AAB20670.1; -.
DR EMBL; AB003791; AAM68399.1; -.
DR PIR; S18643; S18643.
DR PIR; A41639; A41639.
DR FlyBase; FBgn000615; exu.
KW Developmental protein; RNA-binding.
FT DOMAIN 207 246
FT VARIANT 223 223 M -> I.
FT VARIANT 339 339 R -> S (IN P042 MUTANT, LOSS OF EXU

```

FT FUNCTION IN FEMALE.
K -> Q (IN REF. 1).
SQ SEQUENCE 532 AA; 57974 MW; 50BD15B712A62C4E CRC64;
Query Match 2.2%; Score 7; DB 1; Length 532;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
41 STSGSGS 47
| | | | |
Db 209 STSGSGS 215
RESULT 21
CUEO_YERPE STANDARD; PRT; 533 AA.
AC Q82BK0; 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Blue copper oxidase cueo precursor (Copper efflux oxidase).
GN CUEO OR YPO3409.
OS Versinia pestis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Versinia.
OX NCBI_TaxID=632;
[1]
SEQUENCE FROM N.A.
RC STRAIN=CO-92 / Biovar Orientalis;
RX MEDLINE=21470413; PubMed=1158360;
RA Parkhill J., Wren B.W., Thomson N.R., Tlball R.W., Holden M.T.G.,
RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Feltwell T., Hamlin N., Holroyd S., Jagels K., Karpman A.V.,
RA Lammerts W., Skelton J., Stevens K., Whitehead S., Barrett B.G.,
RA "genome sequence of Versinia pestis, the causative agent of plague,";
RL Nature 413:523-527(2001).
CC -1- FUNCTION: Probably involved in periplasmic detoxification of
CC copper by oxidizing Cu(I) to Cu(II) and thus preventing its uptake
CC into the cytoplasm. Possesses phenoloxidase and ferroxidase
CC activities and might be involved in the production of polyphenolic
CC compounds and the prevention of oxidative damage in the periplasm
CC (By similarity).
CC -1- COFACTOR: This protein belongs to the multicopper oxidases which
CC contain three distinct Cu centers known as type 1 or blue, type 2
CC or normal, and type 3 or coupled binuclear (By similarity).
CC -1- SUBUNIT: Monomer (Probable).
CC -1- SUBCELLULAR LOCATION: Periplasmic. It is exported via the Tat
CC pathway (By similarity).
CC -1- INDUCTION: By copper, at increased levels of cytoplasmic cuprous
CC ions (Probable).
CC -1- DOMAIN: The methionine-rich domain could provide binding sites for
CC exogenous copper ions. This methionine-rich region is probably
CC important for copper tolerance in bacteria.
CC -1- SIMILARITY: BELONGS TO THE FAMILY OF MULTICOPPER OXIDASES.
CC -1- SIMILARITY: CONTAINS 3 PLASTOCYANIN-LIKE DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: A41457; CAC92639.1; -
DR InterPro: IPR001117; Cu-oxidase.
DR Pfam: PF00394; Cu-oxidase; 1.
DR PROSITE: PS00080; MULTICOPPER_OXIDASE2; 1.
KW Oxidoreductase; Copper; Metal-binding; Periplasmic; Signal; Repeat;

KW Complete proteome.
FT SIGNAL 1 28
FT CHAIN 29 533
FT DOMAIN 68 164
FT DOMAIN 165 425
FT DOMAIN 426 533
FT DOMAIN 356 415
FT METAL 102 102
FT METAL 104 104
FT METAL 142 142
FT METAL 144 144
FT METAL 458 458
FT METAL 461 461
FT METAL 463 463
FT METAL 514 514
FT METAL 516 516
FT METAL 520 520
SQ SEQUENCE 533 AA; 58328 MW; 5D1570C9B9B1C35 CRC64;
Query Match 2.2%; Score 7; DB 1; Length 533;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
56 LAIGSQV 62
| | | | |
Db 324 LAIGSQV 330
RESULT 22
ERF_MOUSE STANDARD; PRT; 551 AA.
AC P70459;
DT 15-JUN-1998 (Rel. 36, Created)
DT 15-JUN-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ERF-domain transcription factor ERF.
GN ERF.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
SEQUENCE FROM N.A.
RC STRAIN=129/SVJ;
RX MEDLINE=97282708; PubMed=9136988;
RA Liu D., Pavlopoulos E., Modi W., Moschonas N., Mavrothalassitis G.J.;
RT "ERF: genomic organization, chromosomal localization and promoter
RT analysis of the human and mouse genes,";
RL Oncogene 14:1445-1451(1997).
CC -1- FUNCTION: POTENT TRANSCRIPTIONAL REPRESSOR THAT BINDS TO THE H1
CC ELEMENT OF THE ETS2 PROMOTER. MAY REGULATE OTHER GENES INVOLVED
CC IN CELLULAR PROLIFERATION (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC PHOSPHORYLATED BY MULTIPLE KINASES INCLUDING PROBABLY ERK2.
CC -1- SIMILARITY: BELONGS TO THE ETS FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U58533; AAC09474.1; -
DR EMBL: U58534; AAC09474.1; JOINED.
DR HSP: Q01543; 1FLI.
DR MG: MG1.109637; ERF.
DR InterPro: IPR000419; Ets.
DR InterPro: IPR002341; HSF_ETs.

```

DR PFam; PF00178; Ets; 1.
DR PRINTS; PR00454; ETSDOMAIN.
DR SMART; SM00413; ETS; 1.
DR PROSITE; PS00345; ETS_DOMAIN_1; 1.
DR PROSITE; PS00346; ETS_DOMAIN_2; 1.
DR PROSITE; PS00347; ETS_DOMAIN_3; 1.
DR TRANSCRIPTION REGULATION; Repressor; DNA-binding; Nuclear protein;
KW Phosphorylation.
FT DNA_BIND 27 107 ETS-DOMAIN.
FT DOMAIN 166 171 POLY-SER.
FT DOMAIN 290 293 POLY-GLY.
FT DOMAIN 362 373 POLY-SER.
FT DOMAIN 420 425 POLY-PRO.
FT MOD_RES 529 529 PHOSPHORYLATION (BY MAPK1) (BY
SIMILARITY).
SQ SEQUENCE 551 AA; 59050 MW; 5AC1B72FB2743FE5 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 47;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 51 SSGSGLA 57
DB 406 SSGSGLA 412

RESULT 23
FIXN AZOCA STANDARD; PRT; 551 AA.
AC P98056;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cytochrome c oxidase polypeptide I homolog (EC 1.9.3.1).
GN FIXN.
OS Azorhizobium caulinodans.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Hyphomicrobium group; Azorhizobium.
OC NCBI_Taxid=7;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ORS571;
RX MEDLINE=94109675; PubMed=8282187;
RA Mandon K., Kaminski P.A., Mougel C., Desnoues N., Eimerich C.;
RT "Role of the fixXII region of Azorhizobium caulinodans in free-living
and symbiotic nitrogen fixation.";
RL FEMS Microbiol. Lett. 114:185-189(1993).
CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNIT 1-
3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN
CYTOCHROME C OR A QUINOL ARE TRANSFERRED TO THE BINETALLIC CENTER
FORMED BY A HIGH-SPIN HEME AND COPPER B.
CC -1- CATALYTIC ACTIVITY: 4 ferricytochrome c + 2 H(2)O.
CC -1- PATHWAY: Respiratory chain; terminal step.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. CONTAINS 13 OR 14
POTENTIAL TRANSMEMBRANE DOMAINS.
CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC EMBL; X74410; CAA52429.1; -
CC InterPro; IPR000083; COX1.
CC InterPro; IPR004677; CCON.
CC PFam; PF00115; COX1; 1.
CC TIGRfam; TIGR00780; ccon; 1.

```

```

DR PROSITE; PS00077; COX1; 1.
KW Oxidoreductase; Heme; Copper; Transmembrane; Respiratory chain.
FT METAL 132 132 IRON (LOW-SPIN HEME) (BY SIMILARITY).
FT METAL 281 281 COPPER B (BY SIMILARITY).
FT METAL 331 331 COPPER B (BY SIMILARITY).
FT METAL 332 332 COPPER B (BY SIMILARITY).
FT METAL 419 419 IRON (HIGH-SPIN HEME) (BY SIMILARITY).
FT METAL 421 421 IRON (LOW-SPIN HEME) (BY SIMILARITY).
SQ SEQUENCE 551 AA; 61843 MW; 0466F71CD04BFCAD CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 47;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 297 OPSGAL 303
DB 541 OPSGAL 547

RESULT 24
UL25_HSV1 STANDARD; PRT; 580 AA.
AC P10209;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 01-DEC-1992 (Rel. 24, Last annotation update)
DE Varion protein UL25.
GN UL25.
OS Herpes simplex virus (type 1 / strain 17).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Simplexvirus.
OC NCBI_Taxid=10299;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88274327; PubMed=2839594;
RA McGeoch D.J., Dalrymple M.A., Davison A.J., Dolan A., Frame M.C.,
RA McNab D., Perry L.J., Scott J.E., Taylor P.;
RT "The complete DNA sequence of the long unique region in the genome of
herpes simplex virus type 1.";
RL J. Gen. Virol. 69:1531-1574(1988).
CC -1- FUNCTION: VIRION PROTEIN.
CC -1- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL25,
EHV-1 36, EBV BFRF1, HCMV UL77, ILTV ORF2, AND VZV 34.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC EMBL; D10879; BAA01671.1; -
CC EMBL; X14112; CAA32317.1; -
CC PIR; G30084; WMBEWS.
CC InterPro; IPR002493; UL25.
CC PFam; PF01499; UL25; 1.
CC SEQUENCE 580 AA; 62669 MW; 3F8F0B7C122B2536 CRC64;

```

```

Query Match
Best Local Similarity 100.0%; Pred. No. 49;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 191 ATCGSGV 197
DB 167 ATCGSGV 173

RESULT 25
SPFA_ECOLI STANDARD; PRT; 618 AA.
ID SPFA_ECOLI
AC P08395; P77752; Q46723; Q46724; Q46725; Q46726; Q57183;
DT 01-AUG-1988 (Rel. 08, Created)

```

DT 01-AUG-1988 (Rel. 08, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Peptidase IV (EC 3.4.21.-) (Endopeptidase IV) (Signal peptide
 DE peptidase).
 GN SPpA OR B1766.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 NCBI_TaxID=562;
 RX MEDLINE=66250892; PubMed=3522590;
 RA Ichihara S., Suzuki T., Suzuki M., Mizushima S.;
 RT "Molecular cloning and sequencing of the sppA gene and
 RT characterization of the encoded protease IV, a signal peptide
 RT peptidase, of Escherichia coli.";
 RL J. Biol. Chem. 261:9405-9411(1986).
 RN
 RP
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Blythe M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y., Sivasubramanian S., Tagami H., Takada J.,
 RT "The complete genome sequence of Escherichia coli K-12.";
 RT Science 277:1232-1244(1997).
 RN
 RN SEQUENCE FROM N.A.
 RX STRAIN=K12 / MG1655;
 RA Alaba H., Baba T., Fujita K., Hayashi K., Inada T., Itoh T.,
 RA Kawai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
 RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
 RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T., Saito N.,
 RA Sempel G., Seki Y., Sivasubramanian S., Tagami H., Takada J.,
 RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horikuchi T.;
 RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
 RT corresponding to the 28.0-40.1 min region on the linkage map.";
 RL DNA Res. 3:363-377(1996).
 RN
 RN
 RP SEQUENCE OF 110-433 FROM N.A.
 RX STRAIN=Various E. coli strains;
 RX MEDLINE=95064015; PubMed=7973728;
 RA Gutman D.S., Dykhuizen D.E.;
 RT "Clonal divergence in Escherichia coli as a result of recombination,
 RT not mutation.";
 RL Science 266:1380-1383(1994).
 CC -1- FUNCTION: DIGESTION OF THE CLEAVED SIGNAL PEPTIDES. THIS ACTIVITY
 CC IS NECESSARY TO MAINTAIN PROPER SECRETION OF MATURE PROTEINS
 CC ACROSS THE MEMBRANE.
 CC -1- SUBUNIT: HOMOTETRAMER.
 CC (Probable).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
 CC (Probable).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S49.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: M13359; AAA24648.1; -
 CC EMBL: AEO00271; AAC74836.1; -
 CC EMBL: D90820; BAA15557.1; -
 CC EMBL: U13772; AAA57008.1; -
 CC EMBL: U13773; AAA57009.1; -
 CC EMBL: U13774; AAA57010.1; -
 CC EMBL: U13775; AAA57011.1; -
 CC EMBL: U13776; AAA57012.1; -
 CC EMBL: U13777; AAA57013.1; -

DR EMBL: U13778; AAA57014.1; -
 DR EMBL: U13779; AAA57015.1; -
 DR EMBL: U13780; AAA57016.1; -
 DR EMBL: U13782; AAA57017.1; -
 DR EMBL: U13833; AAA57030.1; -
 DR EMBL: U13834; AAA57031.1; -
 DR PIR: A24813; PREC14.
 DR MEROPS: S49.001; -
 DR Ecogene; EG10968; sppA.
 DR Interpro: IPR002142; Peptidase U7.
 DR Interpro: IPR004635; SigPase_SppA36.
 DR Interpro: IPR004634; SigPase_SppA67.
 DR Pfam: PF01343; Peptidase U7; 2.
 DR Prodom: PD002897; Peptidase U7; 1.
 DR TIGRPFAMs: TIGR00705; SppA_67K; 1.
 DR TIGRPFAMs: TIGR00706; SppA_dom; 1.
 KY Hydrolyase; Protease; SppA; -
 FT TRANSMEM 23
 FT TRANSMEM 38 45
 FT TRANSMEM 421 414
 FT TRANSMEM 131 151
 FT TRANSMEM 186 186
 FT TRANSMEM 252 252
 FT TRANSMEM 252 252
 FT TRANSMEM 252 252
 FT TRANSMEM 252 252
 FT TRANSMEM 294 294
 FT TRANSMEM 378 378
 FT TRANSMEM 618 AA; 67233 MW; EBB551C3E2946AC3 CRC64;
 SQ SEQUENCE
 Query Match 2.2%; Score 7; DB 1; Length 618;
 Best local similarity 100.0%; Pred. No. 52;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 21 IFGVIT 27
 Db 434 IFGVIT 440
 RESULT 26
 FATTY YEAST STANDARD; PRT; 623 AA.
 ID P38225;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Probable long-chain fatty acid transport protein.
 GN FATT1 OR YBR041W OR YBR041L.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetaceae; Saccharomycetaceae; Saccharomycetes.
 NCBI_TaxID=4932;
 RN
 RP SEQUENCE FROM N.A.
 RX STRAIN=5288c;
 RA Andre B., Cziepluch C., Hein C., Jauniaux J.C., Urrestarazu A.,
 RA Vissers S.;
 RT Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.
 RN
 RP CHARACTERIZATION.
 RX MEDLINE=97236810; PubMed=9079682;
 RA Faergeman N.J., Dirusso C.C., Elberger A., Knudsen J., Black P.N.;
 RT "Disruption of the Saccharomyces cerevisiae homologue to the murine
 RT fatty acid transport protein impairs uptake and growth on long-chain
 RT fatty acids.";
 RL J. Biol. Chem. 272:8531-8538(1997).
 CC -1- FUNCTION: MAY BE INVOLVED IN LONG-CHAIN FATTY ACIDS UPTAKE, AND
 CC THUS MAY PLAY A PIVOTAL ROLE IN REGULATING THEIR ACCESSIBILITY
 CC PRIOR TO METABOLIC UTILIZATION. MAY PLAY AN IMPORTANT ROLE IN
 CC UPTAKE OF THESE HYDROPHOBIC COMPOUNDS UNDER CONDITIONS WHERE FATTY
 CC ACID SYNTHESIS IS COMPROMISED.

```

CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z35910; CAA84983.1; -.
DR PIR; S45899; S45899.
DR SGD; S0000245; PAT1.
DR InterPro; IPR000873; AMP-bind.
DR Pfam; PF00501; AMP-binding; 1.
DR PROSITE; PS00455; AMP_BINDING; 1.
KW Lipid transport; Transmembrane.
FT TRANSMEM 6 26 POTENTIAL.
FT TRANSMEM 54 71 POTENTIAL.
FT TRANSMEM 149 169 POTENTIAL.
FT TRANSMEM 293 313 POTENTIAL.
FT TRANSMEM 313 313 POTENTIAL.
FT CARBOHYD 184 184 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 289 289 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 534 534 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 591 591 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 623 AA; 71697 MW; 0AE02700B60C8CE CRC64;

Query Match 2.2%; Score 7; DB 1; Length 623;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 186 TEAPFAT 192
Db 398 TEAPFAT 404

RESULT 27
DCMA MOOTH STANDARD; PRT; 729 AA.
ID DCMA MOOTH STANDARD; PRT; 729 AA.
AC P27988;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Carbon monoxide dehydrogenase alpha subunit (EC 1.2.99.2) (COOH).
OS Moorella thermoacetica (Clostridium thermoacetum).
OC Bacteria; Firmicutes; Clostridia; Thermococcobacteriales;
OC Thermococcobacteriaceae; Moorella group; Moorella.
OX NCBI_TaxID=1525;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=92084676; PubMed=1748656;
RA Morton T.A., Runquist J.A., Ragsdale S.W., Shanmugaseundaram T.,
RA Wood H.G., Ljungdahl L.G.;
RT "The primary structure of the subunits of carbon monoxide
RT dehydrogenase/acetyl-CoA synthase from Clostridium thermoacetum.";
RL Biol.Chem. 266:23824-23828(1991).
CC -1- FUNCTION: CATALYZES THE INTERCONVERSION OF CO AND CO2 AND THE
CC SYNTHESIS OF ACETYL-COENZYME A FROM THE METHYLATED CORRIDOID/IRON
CC SULFUR PROTEIN, CO, AND COENZYME A.
CC -1- CATALYTIC ACTIVITY: CO + H(2)O + acceptor = CO(2) + reduced
CC acceptor.
CC -1- COFACTOR: NICKEL.
CC -1- SUBUNIT: Tetramer of two alpha and two beta chains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

CC EMBL; M62727; AAA23229.1; -.
DR PIR; B41670; B41670.
DR InterPro; IPR004461; CdhC.
DR Pfam; PF03598; CdhC; 1.
DR TIGRfam; TIGR00316; cdhC; 1.
KW Oxidoreductase; Nickel.
SQ SEQUENCE 729 AA; 81725 MW; 619BB19D959F5A72 CRC64;

Query Match 2.2%; Score 7; DB 1; Length 729;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 209 PVPYRFG 215
Db 139 PVPYRFG 145

RESULT 28
YKGT CAEEL STANDARD; PRT; 781 AA.
ID YKGT CAEEL STANDARD; PRT; 781 AA.
AC P46557;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Hypothetical 86.0 kDa protein B0285.7 in chromosome III precursor.
GN B0285.7.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Sulston J.;
RA Submitted (JUN-1994) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z34533; CAA84298.1; -.
DR WormPep; B0285.7; CE00646.
DR InterPro; IPR001930; Ala_peptase.
DR Pfam; PF01433; Peptidase_M1; 1.
KW Hypothetical protein; Signal.
FT CHAIN 1 27 POTENTIAL.
FT CHAIN 28 781 HYPOTHETICAL PROTEIN B0285.7.
FT DOMAIN 70 73 POLY-THR.
FT DOMAIN 81 84 POLY-THR.
FT DOMAIN 89 93 POLY-THR.
FT CARBOHYD 54 54 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 183 183 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 341 341 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 375 375 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 520 520 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 781 AA; 85984 MW; A9D5BCD8F567B8B0 CRC64;

Query Match 2.2%; Score 7; DB 1; Length 781;
Best Local Similarity 100.0%; Pred. No. 64;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 SISSGVN 132
Db 257 SISSGVN 263

RESULT 29
CPN_DROME

```

ID CPN DROME STANDARD; PRT; 865 AA.
 AC 002510;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 01-FEB-1994 (Rel. 28, Last annotation update)
 DE Calpocin.
 GN CPN OR CAP.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 OC NCHI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Canton-S;
 RX MEDLINE=93165729; PubMed=8094559;
 RA Martin J.H., Benzer S., Rudnicka M., Miller C.A.;
 RT "Calpocin: a Drosophila photoreceptor cell calcium-binding protein.";
 RL Proc. Natl. Acad. Sci. U.S.A. 90:1531-1535(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Canton-S;
 RX MEDLINE=93165730; PubMed=8434015;
 RA Ballinger D.G., Xue N., Harshman K.D.;
 RT "A Drosophila photoreceptor cell-specific protein, calpocin, binds
 RT calcium and contains a leucine zipper.";
 RL Proc. Natl. Acad. Sci. U.S.A. 90:1536-1540(1993).
 CC -1- FUNCTION: MIGHT FUNCTION AS A CALCIUM-SEQUESTERING "SPONGE" TO
 CC REGULATE THE AMOUNT OF FREE CYTOPLASMIC CALCIUM. IT BINDS 0.3 MOL
 CC OF CA-2 PER MOL OF PROTEIN.
 CC -1- SUBUNIT: HOMODIMER (PROBABLE).
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC; HYPODENSE COMPARTMENT.
 CC -1- TISSUE SPECIFICITY: SOMA AND AXONS OF PHOTORECEPTOR CELLS OF
 CC COMPOUND EYES AND OCCELLI.
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED EARLY IN PHOTORECEPTOR CELL
 CC DEVELOPMENT.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; L02111; AAA28405.1; -;
 CC EMBL; L05080; AAA28420.1; -;
 CC PIR; A47282; A47282.
 CC DR FlyBase; FBgn0010218; Cpn.
 CC KM Calcium-binding.
 CC FT CONFLICT 36 A -> AVAPAVVA (IN REF. 2).
 CC FT CONFLICT 43 I -> T (IN REF. 2).
 CC FT CONFLICT 64 I -> V (IN REF. 2).
 CC FT CONFLICT 76 T -> A (IN REF. 2).
 CC FT CONFLICT 100 P -> PP (IN REF. 2).
 CC FT CONFLICT 126 VQ -> AP (IN REF. 2).
 CC FT CONFLICT 154 I -> V (IN REF. 2).
 CC FT CONFLICT 160 S -> T (IN REF. 2).
 CC FT CONFLICT 160 S -> E (IN REF. 2).
 CC FT CONFLICT 534 S -> E (IN REF. 2).
 CC FT CONFLICT 534 S -> T (IN REF. 2).
 CC FT CONFLICT 699 I -> T (IN REF. 2).
 CC FT CONFLICT 703 V -> L (IN REF. 2).
 CC FT CONFLICT 721 D -> E (IN REF. 2).
 CC SQ SEQUENCE 865 AA; 84781 MW; 2110417E0B0E7CFC CRC64;
 Query Match 2.28; Score 7; DB 1; Length 865;
 Best Local Similarity 100.0%; Pred. No. 70;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 179 APVAPPV 185
 DB 112 APVAPPV 118

RESULT 30
 ID HIRA DROME STANDARD; PRT; 1047 AA.
 AC 017468; O9M303; 046105; 077144;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE HIRA protein homolog (dHRA).
 GN HIRA OR DHM OR CG32153
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 OC NCHI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A. (LONG ISOFORM), AND DEVELOPMENTAL STAGE.
 RC TISSUE=Embryo;
 RX MEDLINE=98278847; PubMed=9611274;
 RA Kirov N., Shitlubs A., Rushlow C.;
 RT "Isolation and characterization of a new gene encoding a member of the
 RT HIRA family of proteins from Drosophila melanogaster.";
 RL Gene 212:323-332(1998).
 RN [2]
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
 RC TISSUE=Embryo;
 RX MEDLINE=98380288; PubMed=9712723;
 RA Llevadot R., Marques G., Pritchard M., Estivill X., Ferrus A.,
 RA Scambler P.;
 RT "Cloning, chromosome mapping and expression analysis of the HIRA gene
 RT from Drosophila melanogaster.";
 RL Biochem. Biophys. Res. Commun. 249:486-491(1998).
 RN [3]
 RP SEQUENCE FROM N.A. (LONG ISOFORM).
 RC STRAIN=Berkley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides F.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.B., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton R.G., Wortman J.R., Yandell M.D., Zheng X., Chen L.X.,
 RA Brannon R.C., Rogers J.-H.C., Blazer R.G., Chang C., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA April J.F., Agbayani A., An H.-U., Andrews-Framnoch C., Baldwin D.,
 RA Balow R.W., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harritis N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegawa C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kemison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasok P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Matei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Markov G., Malshe N.V., Mobarry C., Morris J., Mosnere A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Murty D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon S., Pollard J.D., Puri V., Reese M.G.,
 RA Palazzolo M., Peltman G.S., Pan S., Sanders R.D., Schaefer F., Shen H.,
 RA Reinert K., Remington K., Simpson M., Skupski M.P., Smith T.,
 RA Snie B.C., Sinden-Kiamos I., Simpson M., Strong R., Sun E.,
 RA Spier E., Spradling A.C., Stadelman W., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).


```

CC -1- FUNCTION: COULD PLAY A PART IN MECHANISMS OF TRANSCRIPTIONAL
CC REGULATION SIMILAR TO THAT PLAYED BY YEAST HIR1 AND HIR2 TOGETHER.
CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC SHORT FORM, ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED MATERNALLY AND ZYGOTICALLY
CC THROUGHOUT DEVELOPMENT TO ADULTS (MALE AND FEMALE).
CC -1- SIMILARITY: CONTAINS 4 WD REPEATS (TRP-ASP DOMAINS).
CC -1- SIMILARITY: BELONGS TO THE HIR1 FAMILY OF WD-REPEAT PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF031081; AAC48360.1; -
DR EMBL; AJ222709; CA10954.1; -
DR EMBL; AF071881; AAC64041.1; -
DR EMBL; AE003441; AAF46267.1; -
DR PDBase; PDBgn0022786; Hira.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 7.
DR PRINTS; PR00320; GPROTEINRPT.
DR ProDom; PD000018; WD40; 3.
DR SMART; SM00320; WD40; 6.
DR PROSITE; PS00678; WD_REPEATS_1; 1.
DR PROSITE; PS50082; WD_REPEATS_2; 3.
DR PROSITE; PS50294; WD_REPEATS_REGION; 1.
DR Transcription regulation; Repeat; WD repeat; Nuclear protein;
KM Alternative splicing.
FT REPEAT 68 98 WD 1.
FT REPEAT 127 157 WD 2.
FT REPEAT 170 200 WD 3.
FT REPEAT 264 310 WD 4.
FT DOMAIN 932 942 POLY-THR.
FT DOMAIN 944 972 POLY-SER.
FT VARSPPLIC 63 63 L-> LPVSDKAEFDADVPKML (IN SHORT
FT ISOFORM).
FT VARSPPLIC 430 437 KQGRKIT-> LSLICKIF (IN SHORT ISOFORM).
FT VARSPPLIC 438 1047 MISSING (IN SHORT ISOFORM).
FT CONFLICT 53 53 A-> G (IN REF. 1).
FT CONFLICT 58 58 D-> E (IN REF. 1).
FT CONFLICT 64 64 C-> G (IN REF. 1).
FT CONFLICT 72 72 C-> S (IN REF. 1).
FT CONFLICT 159 163 QAEPH-> RHFN (IN REF. 2; AAC64041).
FT CONFLICT 169 169 K-> E (IN REF. 2; AAC64041).
FT CONFLICT 179 179 S-> W (IN REF. 2; AAC64041).
FT CONFLICT 232 232 G-> A (IN REF. 2; AAC64041).
FT CONFLICT 242 242 N-> D (IN REF. 2; CA10954).
FT CONFLICT 330 330 D-> Y (IN REF. 1).
FT CONFLICT 416 416 A-> V (IN REF. 1).
FT CONFLICT 453 455 MNI-> LNF (IN REF. 1).
FT CONFLICT 459 459 S-> R (IN REF. 2).
FT CONFLICT 536 536 L-> V (IN REF. 1).
FT CONFLICT 1043 1047 KTDQT-> PNRSDMKRKPIPNKPKPL (IN REF. 1).
SQ SEQUENCE 1047 AA; 113415 MW; 3614D5F411DC40C CRC64;

```

Query Match 2.2%; Score 7; DB 1; Length 1047;
 Best Local Similarity 100.0%; Pred. No. 83;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 31
 ATU_STAUU STANDARD; PRT; 1256 AA.
 AC P52081;

```

DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Bifunctional autolysin precursor [includes: N-acetyl-muramoyl-L-alanine
DE amidase (EC 3.5.1.28), Mannosyl-glycoprotein endo-beta-N-
DE acetylglucosamidase (EC 3.2.1.96)].
GN ATL.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxId=1280;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 205-214 AND 776-792.
RC STRAIN=RN450;
RX MEDLINE=95116542; PubMed=7816834;
RA Oshida T., Sugai M., Komatsuza H., Hong Y.-M., Suglnaka H.,
RA Tomasz A.;
RT "A Staphylococcus aureus autolysin that has an N-acetyl-muramoyl-L-
RT alanine amidase domain and an endo-beta-N-acetylglucosaminidase
RT domain: cloning, sequence analysis, and characterization.",
RL Proc. Natl. Acad. Sci. U.S.A. 92:285-289(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=NCITC 8325-4;
RA Foster S.J.;
RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: ENDOHYDROLYSIS OF THE DI-N-ACETYLCHITOSYL UNIT IN
CC HIGH-MANNOSE GLYCOPOLYMERIDES AND GLYCOPOLYMERIDES CONTAINING THE
CC -[MAN]5(GLCNAc)2]-ASN STRUCTURE. ONE N-ACETYL-D-GLUCOSAMINE
CC RESIDUE REMAINS ATTACHED TO THE PROTEIN; THE REST OF THE
CC OLIGOSACCHARIDE IS RELEASED INTACT.
CC -1- CATALYTIC ACTIVITY: Hydrolyzes the link between N-acetyl-muramoyl
CC residues and L-alanine acid residues in certain bacterial cell-wall
CC glycopeptides.
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of the di-N-acetylchitosyl
CC unit in high-mannose glycopeptides and glycoproteins containing
CC the -[Man(GlcNAc)2]Asn-structure. One N-acetyl-D-glucosamine
CC residue remains attached to the protein; the rest of the
CC oligosaccharide is released intact.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- PTM: UNDERGOES PROTEOLYTIC PROCESSING TO GENERATE THE TWO
CC EXTRACELLULAR LYTIC ENZYMES.
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE N-
CC ACETYLMURAMOYL-L-ALANINE AMIDASE FAMILY 2.
CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO FAMILY 73 OF
CC GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D17366; BAA04185.1; -
DR EMBL; L41498; AAA99882.1; -
DR InterPro; IPR002502; Amidase_2.
DR InterPro; IPR002901; Amidase_4.
DR Pfam; PF01510; Amidase_2; 1.
DR Pfam; PF01832; Amidase_4; 1.
DR SMART; SM00047; LY22; 1.
DR Cell wall; Hydrolase; Signal; Multifunctional enzyme; Repeat.
KM CHAIN 1 29 POTENTIAL.
FT CHAIN 30 1256 BIFUNCTIONAL AUTOLYSIN.
FT DOMAIN 199 775 N-ACETYLMURAMOYL-L-ALANINE AMIDASE.
FT DOMAIN 776 1256 ENDO-BETA-N-ACETYLGUCOSAMIDASE.
FT REPEAT 425 589 1.
FT REPEAT 596 758 2.
FT REPEAT 770 932 3.
SQ SEQUENCE 1256 AA; 137384 MW; 2BB76CAA292FDD20 CRC64;

```

Query Match 2.2%; Score 7; DB 1; Length 1256;
 Best Local Similarity 100.0%; Pred. No. 98;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 80 TVSKIAQ 86
 Db 787 TVSKIAQ 793

RESULT 32

CAIE_HUMAN STANDARD; PRT; 1388 AA.

AC P39059; (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Collagen alpha 1(XV) chain precursor.
 GN COL15A1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OX NCBI_TaxID=9606;
 [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Umbilical cord;
 RX MEDLINE=94148920; PubMed=8106446;
 RA Kivirikko S., Heinämäki P., Rehn M.V., Honkanen N., Myers J.C.,
 RA Pihlajaniemi T.;
 RT "Primary structure of the alpha 1 chain of human type XV collagen and
 RT exon-intron organization in the 3' region of the corresponding
 RT gene.";
 RL J. Biol. Chem. 269:4773-4779(1994).
 [2]
 RP SEQUENCE OF 1-569 FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=94140817; PubMed=8307960;
 RA Murgaki Y., Abe N., Nimmiya Y., Olsen B.R., Ooshima A.;
 RT "The human alpha 1(XV) collagen chain contains a large amino-terminal
 RT non-triple helical domain with a tandem repeat structure and homology
 RT to alpha 1(XVIII) collagen.";
 RL J. Biol. Chem. 269:4042-4046(1994).
 [3]
 RP SEQUENCE OF 544-1252 FROM N.A.
 RX MEDLINE=93066196; PubMed=1279671;
 RA Myers J.C., Kivirikko S., Gordon M.K., Dion A.S., Pihlajaniemi T.;
 RT "Identification of a previously unknown human collagen chain, alpha
 RT 1(XV), characterized by extensive interruptions in the triple-helical
 RT region.";
 RL Proc. Natl. Acad. Sci. U.S.A. 89:10144-10148(1992).
 CC -1- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN INTERNAL ORGANS
 CC SUCH AS ADRENAL GLAND, PANCREAS AND KIDNEY.
 CC -1- PPM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC -1- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH
 CC INTERRUPTED HELICES (FACIT) FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR SMART: SM00282; Lang: 1.
 DR SMART: SM00210; TSPN: 1.
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 KW Cell adhesion; Collagen; Glycoprotein; Signal.
 FT SIGNAL 1 25
 FT CHAIN 26 1388
 FT DOMAIN 26 555
 FT DOMAIN 526 573
 FT DOMAIN 574 618
 FT DOMAIN 619 732
 FT DOMAIN 733 763
 FT DOMAIN 764 798
 FT DOMAIN 799 822
 FT DOMAIN 823 867
 FT DOMAIN 868 878
 FT DOMAIN 879 949
 FT DOMAIN 950 983
 FT DOMAIN 1014 1027
 FT DOMAIN 1028 1045
 FT DOMAIN 1046 1052
 FT DOMAIN 1053 1107
 FT DOMAIN 1108 1117
 FT DOMAIN 1118 1132
 FT DOMAIN 1133 1388
 FT REPEAT 358 408
 FT REPEAT 409 459
 FT REPEAT 460 509
 FT REPEAT 510 555
 FT CARBOHYD 306 306
 FT CARBOHYD 324 324
 FT CARBOHYD 687 687
 FT CARBOHYD 807 807
 FT CARBOHYD 814 814
 FT CARBOHYD 1046 1046
 FT CONFLICT 10 10
 FT CONFLICT 49 49
 FT CONFLICT 95 95
 FT CONFLICT 150 150
 FT CONFLICT 204 204
 FT CONFLICT 409 409
 SQ SEQUENCE 1388 AA; 141930 MW; 60822AD925A30930 CRC64;
 Query Match 2.2%; Score 7; DB 1; Length 1388;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 254 MDGASIV 260
 Db 780 MDGASIV 786
 RESULT 33
 ID PMPB_CHIMU STANDARD; PRT; 1672 AA.
 AC Q9P072; (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Probable outer membrane protein pmph precursor (Polymorphic membrane
 DE protein B).
 GN PMPB OR TC0694.
 OS Chlamydia muridarum.
 OC Bacteria; Chlamydiales; Chlamydiaeae; Chlamydia.
 OX NCBI_TaxID=83560;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MoPn / Ni99;
 RX MEDLINE=20150255; PubMed=10684935;
 RA Read T.D., Brunham R.C., Shen C., Shen C., Gyll S.R., Heidelberg J.F.,
 RA White O., Hickey E.K., Peterson J., Uuterkack T., Berry K., Bass S.,
 RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,

RA Gwin M., Nelson W., DeBoy R., Kolney J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.,
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT pneumoniae AR39.";
RT Nucleic Acids Res. 28:1397-1406(2000).
CC -1- SUBCELLULAR LOCATION: CELL WALL SURFACE (ELEMENTARY BODIES)
CC (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE PMP OUTER MEMBRANE PROTEIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE003338; AAF39510.1; -.
CC TIGR; TC0694; -.
CC InterPro; IPR003368; Chlamydia_PMP.
CC InterPro; IPR003357; OMP.
CC Pfam; PF02385; OMP; 1.
CC Pfam; PF02415; DUF145; 3.
CC Outer membrane; Signal; Multigene family; Complete proteome.
CC SIGNAL 1 14 POTENTIAL.
CC CHAIN 15 1672 PROBABLE OUTER MEMBRANE PROTEIN PMPB.
CC SEQUENCE 1672 AA; 176295 MW; 4ABF190DA4DF8BD6 CRC64;

Query Match 2.2%; Score 7; DB 1; Length 1672;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 168 PTPPTPV 174
Db 404 PTPPTPV 410

RESULT 34
Y589 CAEEL STANDARD; PRT; 3178 AA.
AC 009624; 009625; 0969D4;
DT 01-NOV-1995 (Rel. 32, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein ZK945.9 in chromosome II.
GN ZK945.9/ZK945.10.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Rhabditida; Rhabditidae;
OC Rhabditidae; Rhabditinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISCOL N2;
RA Wilkinson-Spratt J.;
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP REVISIONS.
RA Durbin R.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: CONTAINS 1 GPS DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z48544; CAB70192.1; -.
CC EMBL; Z48582; CAB70192.1; JOINED.
CC EMBL; Z48582; CAB70201.1; -.

DR EMBL; Z48544; CAB70201.1; JOINED.
DR WormRep; ZK945.9; CE25697.
DR InterPro; IPR002111; Cat channel TrpL.
DR InterPro; IPR001024; Lipoygenase LH2.
DR InterPro; IPR006336; M-channel nlg.
DR InterPro; IPR002033; PKD_cys_rich.
DR Pfam; PF00520; Ion_trans; 1.
DR Pfam; PF01477; PLAT; 1.
DR Pfam; PF01825; GPS; 1.
DR SMART; SM00303; GPS; 1.
DR SMART; SM00308; LH2; 1.
KW Hypothetical protein; Transmembrane.
FT DOMAIN 266 1196 SER/THR-RICH.
FT DOMAIN 1105 1241 GLY/SER-RICH.
FT DOMAIN 2071 2120 GPS.
FT TRANSMEM 13 30 POTENTIAL.
FT TRANSMEM 51 73 POTENTIAL.
FT TRANSMEM 2139 2161 POTENTIAL.
FT TRANSMEM 2348 2367 POTENTIAL.
FT TRANSMEM 2390 2412 POTENTIAL.
FT TRANSMEM 2451 2468 POTENTIAL.
FT TRANSMEM 2483 2505 POTENTIAL.
FT TRANSMEM 2567 2589 POTENTIAL.
FT TRANSMEM 2836 2858 POTENTIAL.
FT TRANSMEM 2939 2961 POTENTIAL.
FT TRANSMEM 2976 2998 POTENTIAL.
FT TRANSMEM 3038 3060 POTENTIAL.
SQ SEQUENCE 3178 AA; 344726 MW; F8239436D03666CD CRC64;

Query Match 2.2%; Score 7; DB 1; Length 3178;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 41 STSGSGS 47
Db 1206 STSGSGS 1212

RESULT 35
DBHI_BIFLO STANDARD; PRT; 27 AA.
AC P17615;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE DNA-binding protein HBI (Fragment).
OS Bifidobacterium longum.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Bifidobacteriales; Bifidobacteriaceae; Bifidobacterium.
OX NCBI_TaxID=1679;
RN [1]
RP SEQUENCE.
RC STRAIN=ATCC 15703;
RX MEDLINE=90344917; PubMed=2116910;
RA Goshima N., Kano Y., Imamoto F.;
RT "Characterization of HU-like protein from Bifidobacterium longum.";
RL Biochimie 72:207-212(1990).
CC -1- FUNCTION: THIS PROTEIN BELONGS TO THE HISTONE LIKE FAMILY OF
CC PROKARYOTIC DNA-BINDING PROTEINS WHICH ARE CAPABLE OF WRAPPING
CC DNA TO STABILIZE IT, AND PREVENT ITS DENATURATION UNDER EXTREME
CC ENVIRONMENTAL CONDITIONS.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SIMILARITY: BELONGS TO THE BACTERIAL HISTONE-LIKE PROTEIN FAMILY.
CC PIR; A43768; A43768.
CC InterPro; IPR00119; Bac_DNABind.
CC Pfam; PF00216; Bac DNA binding; 1.
DR PROSITE; PS00045; HISTONE LIKE; PARTIAL.
KW DNA-binding; DNA condensation.
FT NON TER 27
SQ SEQUENCE 27 AA; 2863 MW; 6A201A5965A1BD64 CRC64;

Query Match 1.9%; Score 6; DB 1; Length 27;
Best Local Similarity 100.0%; Pred. No. 33;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 VSKIAQ 86
| | | | |
Db 8 VSKIAQ 13

RESULT 36
PRTE ONCMY STANDARD; PRT: 32 AA.

AC P08145; (Rel. 08, Created)
DT 01-AUG-1998 (Rel. 08, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Proteasome TP14.
OS Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
OX NCBI_TaxID=8022;
RN [1]
RP MEDLINE=83272939; PubMed=6308564;
RA Alken J.M., McKenzie D., Zhao H.-Z., States J.C., Dixon G.H.;
RT "Sequence homologies in the proteasome gene family of rainbow trout."
RL Nucleic Acids Res. 11:4907-4922(1983).
CC -1- FUNCTION: PROTEASOMES SUBSTITUTE FOR HISTONES IN THE CHROMATIN OF
CC SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT
CC SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- TISSUE SPECIFICITY: TESTIS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.ibt-sib.ch/announce/>
CC or send an email to license@ibt-sib.ch).
CC -----
CC EMBL: X01595; CAA25748.1; -.
DR PIR: A21211; A21211.
KW Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding;
KW Tests; DNA condensation; Nuclear protein.
FT INIT MET 0
SQ SEQUENCE 32 AA; 4260 MW; 5D388D9D90B796E9 CRC64;

Query Match 1.9%; Score 6; DB 1; Length 32;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 144 SSRPPV 149
| | | | |
Db 6 SSRPPV 11

RESULT 37
YSXC_SULAC STANDARD; PRT: 40 AA.

AC P39477;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 4.5 kDa protein in SOXC 3 region.
OS Sulfolobus acidocaldarius.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=2285;
RN [1]
RP SEQUENCE FROM N.A.
RP STRAIN=ATCC 33909 / NCIB 11770 / DSM 639;
RA MEDLINE=92192013; PubMed=1372250;
RA Luebben M., Kolmerer B., Saraste M.;

RT "An archaeobacterial terminal oxidase combines core structures of two
RT mitochondrial respiratory complexes.";
RL EMO J. 11:805-812(1992).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.ibt-sib.ch/announce/>
CC or send an email to license@ibt-sib.ch).
CC -----
CC EMBL: X62643; CAA44512.1; -.
DR PIR: S21044; S21044.
KW Hypothetical protein.
SQ SEQUENCE 40 AA; 4502 MW; 5EEA6A276C0E9A1D CRC64;

Query Match 1.9%; Score 6; DB 1; Length 40;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 LGIIFG 23
| | | | |
Db 9 LGIIFG 14

RESULT 38
IVBC_NAJNA STANDARD; PRT: 57 AA.

AC P19859;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Venom chymotrypsin inhibitor.
OS Naja naja (Indian cobra).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidodactylus; Squamata; Scleroglossae; Serpentes; Colubroidea;
OC Elapidae; Elapinae; Naja.
OX NCBI_TaxID=35670;
RN [1]
RP SEQUENCE.
RP TISSUE:Venom;
RX MEDLINE=91085579; PubMed=2262001;
RA Shafqat J., Zaidi Z.H., Joernvall H.;
RT "Purification and characterization of a chymotrypsin Kunitz inhibitor
RT type of polypeptide from the venom of cobra (Naja naja naja)."
RL FEBS Lett. 275:6-8(1990).
CC -1- FUNCTION: THIS PROTEIN INHIBITS CHYMOTRYPSIN.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: CONTAINS 1 BPTI/KUNITZ INHIBITOR DOMAIN.
CC PIR: S12957; S12957.
DR HSEB; P00981; IDTK.
DR InterPro: IPR002223; Kunitz BPTI.
DR Pfam: PF00014; Kunitz BPTI; 1.
DR PRINTS: PR00759; BASICPTASE.
DR ProDom: PD000222; Kunitz_BPTI; 1.
DR SMART: SM00131; Ku; 1.
DR PROSITE: PS00280; BPTI_KUNITZ_1;
DR PROSITE: PSS0279; BPTI_KUNITZ_2; 1.
KW Venom; Serine protease inhibitor.
FT DISULFID 5
FT DISULFID 35
FT DISULFID 14
FT DISULFID 38
FT DISULFID 30
FT ACT_SITE 15
SQ SEQUENCE 57 AA; 6508 MW; 4BFB26A010C1737A CRC64;

Query Match 1.9%; Score 6; DB 1; Length 57;
Best Local Similarity 100.0%; Pred. No. 64;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 266 GFVSSY 271
| | | | |
Db 16 GFVSSY 21

```

RESULT 39
YL13_ADE41
ID YL13_ADE41 STANDARD; PRT; 60 AA.
AC P23650;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 6.9 kDa protein in 100 kDa protein region.
OS Human adenovirus type 41.
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
OX NCBI_TaxID=10524;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=Tok;
RX MEDLINE=90272433; PubMed=2349115;
RA Slemenda S.B., Pieniazek N.J., Velarde J. Jr., Pieniazek D.,
RA Luftig R.B.;
RT "Nucleotide sequence of the region coding for 100K and 33K proteins
RT of human enteric adenovirus type 41 (Tok).";
RL Nucleic Acids Res. 18:3069-3069(1990).
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; X52532; CAA36763.1; -
CC PIR; S10210; S10210.
CC Hypothetical protein.
CC SEQUENCE 60 AA; 6897 MW; 9AC2E2D25A196138 CRC64;
SQ
Query Match 1.9%; Score 6; DB 1; Length 60;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 141 AVQSSR 146
Db 21 AVQSSR 26
-----
RESULT 40
EC_MAIZE
ID EC_MAIZE STANDARD; PRT; 76 AA.
AC P43401;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE EC protein homolog (Zinc-metallothionein class II).
OS Zea mays (maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliopsida; Liliopsida; Poales; Poaceae; PACC clade;
OC Panicoideae; Andropogoneae; Zea.
OX NCBI_TaxID=4577;
[1]
RN SEQUENCE FROM N.A.
RP White C.N., Rivin C.J.;
RL Submitted (JUN-1994) to the EMBL/GenBank/DBJ databases.
-!- FUNCTION: BINDS 5 MOLECULES OF ZINC. MAY HAVE A ROLE IN ZN(2+)
-!- HOMEOSTASIS DURING EMBRYOGENESIS (BY SIMILARITY).
-!- SIMILARITY: BELONGS TO THE METALLOTHIONEIN SUPERFAMILY; FAMILY 15.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

CC EMBL; Z34469; CAA84233.1; -
CC EMBL; U10696; AAA19405.1; -
CC MaizeDB; 78896; -
CC InterPro; IPR000316; Metallothio15.
CC Pfam; PF02068; Metallothio_PEC; 1.
CC PRINTS; PR00877; MTPPLANTPEC.
CC Metal-binding; Metal-chiolate cluster; Zinc.
CC INIT MET 0
CC BY SIMILARITY.
SQ SEQUENCE 76 AA; 7574 MW; C9803CBF7F9B1D34 CRC64;
Query Match 1.9%; Score 6; DB 1; Length 76;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 50 TSGSGG 55
Db 21 TSGSGG 26
Search completed: July 8, 2003, 11:19:16
Job time : 43 secs

```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 8, 2003, 11:14:23 ; Search time 80 Seconds
(Without alignments)
829.340 Million cell updates/sec

Title: US-10-018-706-2
Perfect score: 322
Sequence: 1 MVTVAINSQNKPIKRLGL.....LFEFRISNGVYDPTLVK 322

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 671580 seqs, 206047115 residues

Word size : 0

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database :

SPTREMBL_21:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	9	2.8	968	5 Q9VCB2	Q9VCB2 drosophila
2	8	2.5	91	10 Q8S0D0	Q8S0D0 oryza sativ
3	8	2.5	120	10 Q42049	Q42049 arabidopsis
4	8	2.5	164	16 Q34816	Q34816 bacillus su
5	8	2.5	227	16 Q9K644	Q9K644 bacillus ha
6	8	2.5	244	3 Q05905	Q05905 saccharomyc
7	8	2.5	306	10 Q9SET1	Q9SET1 arabidopsis
8	8	2.5	334	10 Q9LIE9	Q9LIE9 arabidopsis
9	8	2.5	376	10 Q39353	Q39353 brassica na
10	8	2.2	88	4 Q9UC05	Q9UC05 homo sapien
11	7	2.2	101	16 Q9CDX3	Q9CDX3 lactococcus
12	7	2.2	111	10 Q8S142	Q8S142 oryza sativ
13	7	2.2	122	10 Q8RUE2	Q8RUE2 oryza sativ
14	7	2.2	133	10 Q9ZWI6	Q9ZWI6 zea mays (m
15	7	2.2	133	10 Q9ST25	Q9ST25 zea mays (m
16	7	2.2	136	10 Q9FWP4	Q9FWP4 oryza sativ

17	7	2.2	141	5 P91181	P91181 caenorhabdi
18	7	2.2	142	10 Q9FWP0	Q9FWP0 oryza sativ
19	7	2.2	165	17 Q8TFB9	Q8TFB9 methanosc
20	7	2.2	176	16 Q926C3	Q926C3 rhizobium m
21	7	2.2	177	5 Q9XW11	Q9XW11 caenorhabdi
22	7	2.2	179	16 Q99YU1	Q99YU1 streptococ
23	7	2.2	180	6 Q95KR4	Q95KR4 sus scrofa
24	7	2.2	189	16 Q8ZFP8	Q8ZFP8 yersinia pe
25	7	2.2	195	5 Q9VJ78	Q9VJ78 drosophila
26	7	2.2	205	16 Q8XXR6	Q8XXR6 ralsstonia s
27	7	2.2	208	10 Q81130	Q81130 arabidopsis
28	7	2.2	208	10 Q9SFD8	Q9SFD8 arabidopsis
29	7	2.2	211	16 Q9S2D3	Q9S2D3 streptococ
30	7	2.2	212	10 Q9FYQ4	Q9FYQ4 oryza sativ
31	7	2.2	213	5 Q9VY29	Q9VY29 drosophila
32	7	2.2	216	8 Q8SN98	Q8SN98 euglena gra
33	7	2.2	225	17 Q92808	Q92808 archaeoglob
34	7	2.2	226	16 Q92KU4	Q92KU4 rhizobium m
35	7	2.2	228	4 Q9H5P8	Q9H5P8 homo sapien
36	7	2.2	241	16 Q9KLV7	Q9KLV7 vibrio chol
37	7	2.2	241	17 Q8RTU2	Q8RTU2 methanopyru
38	7	2.2	249	17 Q9HNY3	Q9HNY3 halobacteri
39	7	2.2	259	16 Q51373	Q51373 borrelia bu
40	7	2.2	269	5 Q94238	Q94238 caenorhabdi
41	7	2.2	276	10 Q9SPD9	Q9SPD9 zea mays (m
42	7	2.2	281	16 Q9SPD8	Q9SPD8 zea mays (m
43	7	2.2	281	16 Q9EMD9	Q9EMD9 streptococ
44	7	2.2	285	17 Q2S731	Q2S731 methanobact
45	7	2.2	288	2 Q52246	Q52246 enterococu
46	7	2.2	291	16 Q9S5F0	Q9S5F0 rhizobium 1
47	7	2.2	302	12 Q91HNS	Q91HNS human polio
48	7	2.2	303	10 Q41814	Q41814 zea mays (m
49	7	2.2	308	16 Q98L55	Q98L55 rhizobium 1
50	7	2.2	309	16 Q8Y158	Q8Y158 brucella me
51	7	2.2	310	2 Q51853	Q51853 pseudomonas
52	7	2.2	312	2 Q9K4R9	Q9K4R9 pseudomonas
53	7	2.2	312	11 Q8VGC0	Q8VGC0 mus musculu
54	7	2.2	326	4 Q62449	Q62449 caenorhabdi
55	7	2.2	328	4 Q96E29	Q96E29 homo sapien
56	7	2.2	328	4 Q8WYX0	Q8WYX0 homo sapien
57	7	2.2	328	10 Q42366	Q42366 zea mays (m
58	7	2.2	329	3 Q74489	Q74489 schizosacch
59	7	2.2	334	11 Q9D205	Q9D205 mus musculu
60	7	2.2	339	5 Q18567	Q18567 caenorhabdi
61	7	2.2	344	2 Q82858	Q82858 acrobacter
62	7	2.2	350	10 Q41719	Q41719 zea diptope
63	7	2.2	351	16 Q8ZD12	Q8ZD12 yersinia pe
64	7	2.2	355	16 Q9ZM09	Q9ZM09 helicobacte
65	7	2.2	361	5 Q9W1P2	Q9W1P2 drosophila
66	7	2.2	368	5 Q9XUK0	Q9XUK0 caenorhabdi
67	7	2.2	369	10 Q40692	Q40692 oryza sativ
68	7	2.2	374	5 Q9N1H3	Q9N1H3 drosophila
69	7	2.2	377	16 Q97D06	Q97D06 clostridium
70	7	2.2	380	16 Q98180	Q98180 rhizobium 1
71	7	2.2	381	5 Q94399	Q94399 caenorhabdi
72	7	2.2	387	5 Q16316	Q16316 caenorhabdi
73	7	2.2	387	16 Q8YFP9	Q8YFP9 anabaena sp
74	7	2.2	388	17 Q97B14	Q97B14 thermoplasma
75	7	2.2	389	2 Q9X9P9	Q9X9P9 streptococ
76	7	2.2	390	5 Q9GRY4	Q9GRY4 caenorhabdi
77	7	2.2	391	16 Q9PD30	Q9PD30 xylella fas
78	7	2.2	392	5 Q9W5S7	Q9W5S7 caenorhabdi
79	7	2.2	393	5 Q9N2Z3	Q9N2Z3 caenorhabdi
80	7	2.2	394	5 Q9GNA3	Q9GNA3 drosophila
81	7	2.2	395	5 Q9GNW2	Q9GNW2 drosophila
82	7	2.2	395	5 Q9V181	Q9V181 drosophila
83	7	2.2	395	5 Q9U4F5	Q9U4F5 drosophila
84	7	2.2	395	5 Q9UGN4	Q9UGN4 drosophila
85	7	2.2	396	5 Q44435	Q44435 drosophila
86	7	2.2	396	5 Q9W2S5	Q9W2S5 drosophila
87	7	2.2	400	2 Q938U4	Q938U4 streptococ
88	7	2.2	400	16 Q9A0M2	Q9A0M2 streptococ
89	7	2.2	409	5 Q9GND0	Q9GND0 drosophila

90	7	2	2	409	16	088012	088012 streptomyces	163	7	2	2	450	8	095R03	095bq3 trichocentr
91	7	2	2	410	5	09NH9	09nh9 drosophila	164	7	2	2	450	8	095R02	095b2 trichocentr
92	7	2	2	410	5	09GN46	09gn46 drosophila	165	7	2	2	450	8	095R01	095b0 trichocentr
93	7	2	2	410	5	09TW27	09tw27 drosophila	166	7	2	2	450	8	095R00	095b0 trichocentr
94	7	2	2	414	10	09M0F6	09m0f6 arabisdopsis	167	7	2	2	450	8	095R09	095p9 trichopilla
95	7	2	2	416	10	09M0F6	09m0f6 arabisdopsis	168	7	2	2	450	8	095R08	095p8 trichopilla
96	7	2	2	421	10	092Y19	092y19 arabidopsis	169	7	2	2	450	8	095R07	095p7 trichopilla
97	7	2	2	421	10	092Y19	092y19 arabidopsis	170	7	2	2	451	8	095R06	095p6 houletia t
98	7	2	2	422	10	092Y19	092y19 arabidopsis	171	7	2	2	451	8	095R05	095p5 houletia t
99	7	2	2	424	16	092ZD8	092zd8 rhabdium m	172	7	2	2	451	8	095R04	095p4 houletia t
100	7	2	2	429	16	08TKV2	08tkv2 anabaena sp	173	7	2	2	451	8	095R03	095p3 houletia t
101	7	2	2	431	10	08S6Y5	08s6y5 oryza sativ	174	7	2	2	451	8	095R02	095p2 houletia t
102	7	2	2	432	2	092A36	092a36 streptomyces	175	7	2	2	452	8	095R01	095p1 houletia t
103	7	2	2	440	8	09BA39	09ba39 dictyaea mur	176	7	2	2	452	8	095R00	095p0 houletia t
104	7	2	2	441	8	09BAW4	09baw4 lockhartia	177	7	2	2	452	8	095R00	095p0 houletia t
105	7	2	2	441	8	09BAW3	09baw3 lockhartia	178	7	2	2	453	8	095B01	095b1 houletia t
106	7	2	2	442	8	09BAW3	09baw3 lockhartia	179	7	2	2	453	8	095B02	095b2 houletia t
107	7	2	2	442	8	09BAW3	09baw3 lockhartia	180	7	2	2	453	8	095B03	095b3 houletia t
108	7	2	2	445	8	09BAW3	09baw3 lockhartia	181	7	2	2	454	8	095B04	095b4 houletia t
109	7	2	2	446	8	09BAW3	09baw3 lockhartia	182	7	2	2	455	8	095B05	095b5 houletia t
110	7	2	2	447	8	09BAW3	09baw3 lockhartia	183	7	2	2	456	8	095B06	095b6 houletia t
111	7	2	2	447	8	09BAW3	09baw3 lockhartia	184	7	2	2	457	8	095B07	095b7 houletia t
112	7	2	2	448	3	096U22	096u22 neurospora	185	7	2	2	457	8	095B08	095b8 houletia t
113	7	2	2	448	8	08WIE1	08wie1 caucasia pha	186	7	2	2	459	8	095B09	095b9 houletia t
114	7	2	2	448	8	08WIE1	08wie1 caucasia pha	187	7	2	2	459	8	095B10	095b10 houletia t
115	7	2	2	448	8	08WIE1	08wie1 caucasia pha	188	7	2	2	462	8	095B11	095b11 houletia t
116	7	2	2	449	2	056769	056769 xanthomonas	189	7	2	2	462	8	095B12	095b12 houletia t
117	7	2	2	449	8	095B09	095b09 cyrtocochium	190	7	2	2	464	8	095B13	095b13 houletia t
118	7	2	2	449	8	08WIE3	08wie3 miltoniopsi	191	7	2	2	466	8	095B14	095b14 houletia t
119	7	2	2	449	8	08WIE3	08wie3 miltoniopsi	192	7	2	2	467	8	095B15	095b15 houletia t
120	7	2	2	449	8	08WIE3	08wie3 miltoniopsi	193	7	2	2	467	8	095B16	095b16 houletia t
121	7	2	2	449	8	08WIE3	08wie3 miltoniopsi	194	7	2	2	467	8	095B17	095b17 houletia t
122	7	2	2	449	8	08WIE3	08wie3 miltoniopsi	195	7	2	2	468	8	095B18	095b18 houletia t
123	7	2	2	449	8	08WIE3	08wie3 miltoniopsi	196	7	2	2	468	8	095B19	095b19 houletia t
124	7	2	2	449	8	08WIE3	08wie3 miltoniopsi	197	7	2	2	469	8	095B20	095b20 houletia t
125	7	2	2	449	8	08WIE3	08wie3 miltoniopsi	198	7	2	2	469	8	095B21	095b21 houletia t
126	7	2	2	449	8	08WIE3	08wie3 miltoniopsi	199	7	2	2	470	8	095B22	095b22 houletia t
127	7	2	2	449	8	08WIE3	08wie3 miltoniopsi	200	7	2	2	470	8	095B23	095b23 houletia t
128	7	2	2	450	8	095B08	095b08 cyrtocochium	201	7	2	2	471	8	095B24	095b24 houletia t
129	7	2	2	450	8	095B08	095b08 cyrtocochium	202	7	2	2	471	8	095B25	095b25 houletia t
130	7	2	2	450	8	095B08	095b08 cyrtocochium	203	7	2	2	471	8	095B26	095b26 houletia t
131	7	2	2	450	8	095B08	095b08 cyrtocochium	204	7	2	2	472	8	095B27	095b27 houletia t
132	7	2	2	450	8	095B08	095b08 cyrtocochium	205	7	2	2	472	8	095B28	095b28 houletia t
133	7	2	2	450	8	095B08	095b08 cyrtocochium	206	7	2	2	472	8	095B29	095b29 houletia t
134	7	2	2	450	8	095B08	095b08 cyrtocochium	207	7	2	2	473	8	095B30	095b30 houletia t
135	7	2	2	450	8	095B08	095b08 cyrtocochium	208	7	2	2	473	8	095B31	095b31 houletia t
136	7	2	2	450	8	095B08	095b08 cyrtocochium	209	7	2	2	474	8	095B32	095b32 houletia t
137	7	2	2	450	8	095B08	095b08 cyrtocochium	210	7	2	2	474	8	095B33	095b33 houletia t
138	7	2	2	450	8	095B08	095b08 cyrtocochium	211	7	2	2	481	2	032391	032391 staphylococ
139	7	2	2	450	8	095B08	095b08 cyrtocochium	212	7	2	2	484	8	095B34	095b34 houletia t
140	7	2	2	450	8	095B08	095b08 cyrtocochium	213	7	2	2	486	8	095B35	095b35 houletia t
141	7	2	2	450	8	095B08	095b08 cyrtocochium	214	7	2	2	489	8	095B36	095b36 houletia t
142	7	2	2	450	8	095B08	095b08 cyrtocochium	215	7	2	2	491	8	095B37	095b37 houletia t
143	7	2	2	450	8	095B08	095b08 cyrtocochium	216	7	2	2	492	8	095B38	095b38 houletia t
144	7	2	2	450	8	095B08	095b08 cyrtocochium	217	7	2	2	497	8	08WJ54	08wj54 myoxanthus
145	7	2	2	450	8	095B08	095b08 cyrtocochium	218	7	2	2	498	8	08WJ54	08wj54 myoxanthus
146	7	2	2	450	8	095B08	095b08 cyrtocochium	219	7	2	2	498	13	091882	091882 xenopus lae
147	7	2	2	450	8	095B08	095b08 cyrtocochium	220	7	2	2	499	2	09KXPS	09kxps listeria mo
148	7	2	2	450	8	095B08	095b08 cyrtocochium	221	7	2	2	499	16	09KXPS	09kxps listeria mo
149	7	2	2	450	8	095B08	095b08 cyrtocochium	222	7	2	2	502	11	035817	035817 rattus norv
150	7	2	2	450	8	095B08	095b08 cyrtocochium	223	7	2	2	503	8	095B30	095b30 houletia t
151	7	2	2	450	8	095B08	095b08 cyrtocochium	224	7	2	2	503	8	095B31	095b31 houletia t
152	7	2	2	450	8	095B08	095b08 cyrtocochium	225	7	2	2	505	8	095B32	095b32 houletia t
153	7	2	2	450	8	095B08	095b08 cyrtocochium	226	7	2	2	505	11	061271	061271 mus musculo
154	7	2	2	450	8	095B08	095b08 cyrtocochium	227	7	2	2	505	13	095B33	095b33 houletia t
155	7	2	2	450	8	095B08	095b08 cyrtocochium	228	7	2	2	506	13	095B34	095b34 houletia t
156	7	2	2	450	8	095B08	095b08 cyrtocochium	229	7	2	2	508	5	095B35	095b35 houletia t
157	7	2	2	450	8	095B08	095b08 cyrtocochium	230	7	2	2	508	5	095B36	095b36 houletia t
158	7	2	2	450	8	095B08	095b08 cyrtocochium	231	7	2	2	509	8	095B37	095b37 houletia t
159	7	2	2	450	8	095B08	095b08 cyrtocochium	232	7	2	2	509	8	095B38	095b38 houletia t
160	7	2	2	450	8	095B08	095b08 cyrtocochium	233	7	2	2	510	8	095B39	095b39 houletia t
161	7	2	2	450	8	095B08	095b08 cyrtocochium	234	7	2	2	510	8	095B40	095b40 houletia t
162	7	2	2	450	8	095B08	095b08 cyrtocochium	235	7	2	2	510	8	095B41	095b41 houletia t

236	7	2.2	510	8	Q9BB09	Q9b09 stanhopaea e	309	7	2.2	842	5	Q9VGC9	Q9VGC9 drosophila
237	7	2.2	510	8	Q9BB08	Q9b08 stanhopaea w	310	7	2.2	843	13	Q8QFN9	Q8QFN9 gallus galli
238	7	2.2	510	8	Q9BB07	Q9b07 stanhopaea s	311	7	2.2	852	5	Q9VGC0	Q9VGC0 drosophila
239	7	2.2	510	8	Q9BB06	Q9b06 stanhopaea t	312	7	2.2	855	5	Q8T9D8	Q8T9D8 drosophila
240	7	2.2	510	8	Q9BB05	Q9b05 stanhopaea r	313	7	2.2	857	16	Q98IK2	Q98IK2 rhizobium 1
241	7	2.2	510	8	Q9BB04	Q9b04 stanhopaea a	314	7	2.2	864	5	Q95U45	Q95U45 drosophila
242	7	2.2	510	8	Q9BB02	Q9b02 stievekingia	315	7	2.2	864	5	Q9VGC8	Q9VGC8 drosophila
243	7	2.2	510	8	Q9BB01	Q9b01 stievekingia	316	7	2.2	886	2	Q91CV5	Q91CV5 rhodospirillum
244	7	2.2	510	8	Q9BB00	Q9b00 embreea rod	317	7	2.2	887	16	Q8WY17	Q8WY17 anabaena sp
245	7	2.2	510	8	Q9BAZ9	Q9ba9 coryantes	318	7	2.2	900	5	Q9NTG9	Q9NTG9 drosophila
246	7	2.2	510	8	Q9BAZ8	Q9ba8 coryantes	319	7	2.2	929	5	Q9NCG5	Q9NCG5 drosophila
247	7	2.2	510	8	Q9BAZ2	Q9ba2 schlimmia s	320	7	2.2	943	16	Q93IU4	Q93IU4 staphylococcus
248	7	2.2	510	8	Q9BAZ1	Q9ba1 schlimmia s	321	7	2.2	945	5	Q9NKT7	Q9NKT7 leishmania
249	7	2.2	510	8	Q9BAZ0	Q9ba0 stanhopaea c	322	7	2.2	955	5	Q9VW55	Q9VW55 drosophila
250	7	2.2	510	8	Q9BAZ8	Q9ba8 braemia vit	323	7	2.2	979	5	Q95ZC3	Q95ZC3 leishmania
251	7	2.2	510	8	Q9BAU9	Q9ba9 eriopsis bl	324	7	2.2	1035	12	Q91TP4	Q91TP4 tupiaa hep
252	7	2.2	512	8	Q8WJS2	Q8wj2 restreptia a	325	7	2.2	1062	5	Q9VTV2	Q9VTV2 drosophila
253	7	2.2	512	8	Q8WJM8	Q8wj8 pleurothall	326	7	2.2	1119	16	Q8YR12	Q8YR12 anabaena sp
254	7	2.2	513	3	Q9Y7P3	Q9y7p3 schizosacch	327	7	2.2	1154	11	Q91ZS0	Q91ZS0 mus musculus
255	7	2.2	513	8	Q8WJT2	Q8wj2 pleurothall	328	7	2.2	1172	10	Q91LMY9	Q91LMY9 oryza sativ
256	7	2.2	513	8	Q8WJT1	Q8wj1 pleurothall	329	7	2.2	1208	5	Q97101	Q97101 dictyostelium
257	7	2.2	516	8	Q8WJS9	Q8wj9 frondaria c	330	7	2.2	1227	5	Q95ZT4	Q95ZT4 leishmania
258	7	2.2	516	8	Q8WJS0	Q8wj0 octomeria g	331	7	2.2	1236	4	Q9C0I2	Q9C0I2 homo sapien
259	7	2.2	516	16	Q9F3B2	Q9f3b2 streptomycete	332	7	2.2	1258	16	Q93V41	Q93V41 staphylococcus
260	7	2.2	517	8	Q8WJV2	Q8wj2 dracula chl	333	7	2.2	1253	5	Q8WQ61	Q8WQ61 drosophila
261	7	2.2	517	8	Q8WJV0	Q8wj0 masdevallia	334	7	2.2	1305	16	Q8YFB6	Q8YFB6 bruceella me
262	7	2.2	517	8	Q8WJU8	Q8wj8 trietella	335	7	2.2	1336	10	Q81074	Q81074 arabiopsis
263	7	2.2	517	8	Q8WJU4	Q8wj4 dryadella e	336	7	2.2	1388	4	Q9Y4W4	Q9Y4W4 homo sapien
264	7	2.2	517	8	Q8WJU1	Q8wj1 scaphosepal	337	7	2.2	1428	5	Q94341	Q94341 halictus ru
265	7	2.2	517	8	Q8WJU0	Q8wj0 pleurothall	338	7	2.2	1621	5	Q8T6J5	Q8T6J5 dictyostelium
266	7	2.2	517	8	Q8WJ79	Q8wj79 condylago r	339	7	2.2	1628	4	Q9HCL7	Q9HCL7 homo sapien
267	7	2.2	517	8	Q8WJ78	Q8wj78 pleurothall	340	7	2.2	1847	5	Q9NKL5	Q9NKL5 leishmania
268	7	2.2	517	8	Q8WJ77	Q8wj77 pleurothall	341	7	2.2	2048	16	Q8Y1I8	Q8Y1I8 raletonia s
269	7	2.2	517	8	Q8WJ74	Q8wj74 pleurothall	342	7	2.2	2160	3	Q13488	Q13488 magnaporthe
270	7	2.2	517	8	Q8WJ73	Q8wj73 pleurothall	343	7	2.2	2160	3	Q13328	Q13328 magnaporthe
271	7	2.2	517	8	Q8WJ70	Q8wj70 platysteie	344	7	2.2	2238	11	Q91ZS1	Q91ZS1 mus musculus
272	7	2.2	517	8	Q8WJS7	Q8wj7 trichosalepi	345	7	2.2	2241	12	Q91JH2	Q91JH2 avian param
273	7	2.2	517	8	Q8WJS6	Q8wj6 dresaleirell	346	7	2.2	2271	5	Q62178	Q62178 caenorhabditis
274	7	2.2	517	8	Q8WJS5	Q8wj5 myoxanthus	347	7	2.2	2271	16	Q99QY4	Q99QY4 staphylococcus
275	7	2.2	517	8	Q8WJS3	Q8wj3 restreptiops	348	7	2.2	2279	5	Q93529	Q93529 caenorhabditis
276	7	2.2	517	8	Q8WJS1	Q8wj1 restreptiell	349	7	2.2	2283	2	Q8VQ99	Q8VQ99 staphylococcus
277	7	2.2	517	8	Q8WJR8	Q8wj8 lepanthopsl	350	7	2.2	2406	4	Q98ZS0	Q98ZS0 homo sapien
278	7	2.2	517	8	Q8WJM7	Q8wj7 pleurothall	351	7	2.2	2780	11	Q91ZS2	Q91ZS2 mus musculus
279	7	2.2	519	8	Q8WJV1	Q8wj1 masdevallia	352	7	2.2	2786	2	Q48926	Q48926 mycobacterium
280	7	2.2	519	8	Q8WJU7	Q8wj7 luarella pe	353	7	2.2	3010	12	Q81825	Q81825 hepatitis c
281	7	2.2	520	8	Q8WJR6	Q8wj6 pleurothall	354	7	2.2	3010	12	Q68Z85	Q68Z85 hepatitis c
282	7	2.2	527	8	Q8WJU2	Q8wj2 pleurothall	355	7	2.2	3063	12	Q8QXN3	Q8QXN3 sugarcane m
283	7	2.2	535	8	Q8WJU5	Q8wj5 dryadella s	356	7	2.2	3063	12	Q8QXN2	Q8QXN2 sugarcane m
284	7	2.2	542	5	Q23446	Q23446 caenorhabditis	357	7	2.2	3064	12	Q8QXN4	Q8QXN4 sugarcane m
285	7	2.2	556	3	Q9HG18	Q9hg18 candida dub	358	7	2.2	3069	16	P95029	P95029 mycobacterium
286	7	2.2	554	16	Q9RDP9	Q9rd9 streptomycete	359	7	2.2	3076	16	Q9X7E2	Q9X7E2 mycobacterium
287	7	2.2	571	5	Q9VKC6	Q9vk6 drosophila	360	7	2.2	3178	5	Q963D4	Q963D4 caenorhabditis
288	7	2.2	571	5	Q8T0J2	Q8t0j2 drosophila	361	7	2.2	6298	11	Q8VHN7	Q8VHN7 mus musculus
289	7	2.2	586	16	Q98J75	Q98j75 rhizobium 1	362	7	2.2	14	11	Q920G5	Q920G5 mus musculus
290	7	2.2	603	16	Q9X1Q9	Q9x1q9 thermocoga	363	7	2.2	26	12	Q9QRT9	Q9QRT9 hepatitis c
291	7	2.2	618	16	Q8XDM5	Q8xdm5 escherichia	364	7	2.2	39	4	Q9UDS9	Q9UDS9 homo sapien
292	7	2.2	634	5	Q17475	Q17474 caenorhabditis	365	7	2.2	39	5	Q9UD56	Q9UD56 homo sapien
293	7	2.2	634	5	Q17474	Q17474 caenorhabditis	366	7	2.2	44	2	Q9F283	Q9F283 parametecium
294	7	2.2	634	10	Q8S9A9	Q8s9a9 paeisiflora	367	7	2.2	57	10	Q23574	Q23574 arabiopsis
295	7	2.2	635	10	Q49153	Q49153 pium sativ	368	7	2.2	51	13	Q81148	Q81148 notophthalm
296	7	2.2	635	10	Q65871	Q65871 pium sativ	369	7	2.2	54	11	Q8VE24	Q8VE24 mus musculus
297	7	2.2	636	10	Q9ZS00	Q9zsg0 phaseolus a	370	7	2.2	55	10	Q91X03	Q91X03 oryza sativ
298	7	2.2	652	16	Q9ZG62	Q9zsg2 rickettsia	371	7	2.2	56	10	Q24546	Q24546 victoria regia
299	7	2.2	661	5	Q9NDZ6	Q9ndz6 leishmania	372	7	2.2	57	4	Q13374	Q13374 homo sapien
300	7	2.2	669	5	Q60021	Q60021 saccharomyc	373	7	2.2	57	10	Q23574	Q23574 arabiopsis
301	7	2.2	731	5	Q21138	Q21138 caenorhabditis	374	7	2.2	57	16	Q8ZBC9	Q8ZBC9 yerisina pe
302	7	2.2	767	2	Q935Y5	Q935y5 synecococc	375	7	2.2	58	12	Q90680	Q90680 human herpe
303	7	2.2	772	17	Q51963	Q51963 halobacteri	376	7	2.2	59	12	Q90676	Q90676 human herpe
304	7	2.2	789	5	Q9U2M6	Q9u2m6 caenorhabditis	377	7	2.2	59	12	Q90684	Q90684 human herpe
305	7	2.2	814	10	Q91X50	Q91x50 arabiopsis	378	7	2.2	59	12	Q90685	Q90685 human herpe
306	7	2.2	836	5	Q61899	Q61899 caenorhabditis	379	7	2.2	59	12	Q90690	Q90690 human herpe
307	7	2.2	836	5	Q9VB05	Q9vb05 drosophila	380	7	2.2	61	11	Q95SC9	Q95SC9 mus musculus
308	7	2.2	840	10	Q9LPP2	Q9lpp2 arabidopsis	381	7	2.2	62	16	Q98H04	Q98H04 rhizobium 1

382	6	1.9	65	2	Q9RCZ0	Q9fcd0 erwinia ste	455	6	1.9	110	12	Q9YRV8	Q9YRV8 rice tungro
383	6	1.9	69	2	P77762	P77762 escherichia	456	6	1.9	113	12	Q9ANP8	Q9ANP8 bradyrhizob
384	6	1.9	69	12	Q90677	Q90677 human herpes	457	6	1.9	113	11	Q9D0X6	Q9D0X6 mus musculus
385	6	1.9	69	12	Q90678	Q90678 human herpes	458	6	1.9	113	12	Q9D0X6	Q9D0X6 mus musculus
386	6	1.9	69	12	Q90679	Q90679 human herpes	459	6	1.9	113	11	Q9D0X6	Q9D0X6 mus musculus
387	6	1.9	69	12	Q90681	Q90681 human herpes	460	6	1.9	113	14	Q960R4	Q960R4 sulfobus
388	6	1.9	69	12	Q90682	Q90682 human herpes	461	6	1.9	113	17	Q960R4	Q960R4 sulfobus
389	6	1.9	69	12	Q90683	Q90683 human herpes	462	6	1.9	113	15	Q960R4	Q960R4 sulfobus
390	6	1.9	69	12	Q90686	Q90686 human herpes	463	6	1.9	115	16	Q960R4	Q960R4 sulfobus
391	6	1.9	69	12	Q90687	Q90687 human herpes	464	6	1.9	115	17	Q960R4	Q960R4 sulfobus
392	6	1.9	69	12	Q90688	Q90688 human herpes	465	6	1.9	117	10	Q960R4	Q960R4 sulfobus
393	6	1.9	69	12	Q90689	Q90689 human herpes	466	6	1.9	118	11	Q960R4	Q960R4 sulfobus
394	6	1.9	69	12	Q90691	Q90691 human herpes	467	6	1.9	118	11	Q960R4	Q960R4 sulfobus
395	6	1.9	69	12	Q90692	Q90692 human herpes	468	6	1.9	119	16	Q960R4	Q960R4 sulfobus
396	6	1.9	71	12	Q90AN7	Q90AN7 human herpes	469	6	1.9	119	16	Q960R4	Q960R4 sulfobus
397	6	1.9	72	5	Q8T2G0	Q8T2G0 dictyostell	470	6	1.9	121	10	Q960R4	Q960R4 sulfobus
398	6	1.9	76	11	Q9J180	Q9J180 mus musculu	471	6	1.9	121	17	Q960R4	Q960R4 sulfobus
399	6	1.9	76	12	Q8VAF7	Q8VAF7 white spot	472	6	1.9	123	5	Q960R4	Q960R4 sulfobus
400	6	1.9	78	2	Q9WVR9	Q9WVR9 pseudomonas	473	6	1.9	123	16	Q960R4	Q960R4 sulfobus
401	6	1.9	79	5	Q9BKK5	Q9BKK5 pseudomonas	474	6	1.9	126	3	Q960R4	Q960R4 sulfobus
402	6	1.9	79	5	Q9BKK5	Q9BKK5 pseudomonas	475	6	1.9	126	5	Q960R4	Q960R4 sulfobus
403	6	1.9	81	5	Q9VZP9	Q9VZP9 pseudomonas	476	6	1.9	127	10	Q960R4	Q960R4 sulfobus
404	6	1.9	81	12	Q90AN8	Q90AN8 human herpes	477	6	1.9	128	9	Q960R4	Q960R4 sulfobus
405	6	1.9	81	12	Q90AN8	Q90AN8 human herpes	478	6	1.9	128	17	Q960R4	Q960R4 sulfobus
406	6	1.9	81	12	Q90AN5	Q90AN5 human herpes	479	6	1.9	129	5	Q960R4	Q960R4 sulfobus
407	6	1.9	81	12	Q90AN5	Q90AN5 human herpes	480	6	1.9	129	10	Q960R4	Q960R4 sulfobus
408	6	1.9	81	12	Q90AN5	Q90AN5 human herpes	481	6	1.9	129	12	Q960R4	Q960R4 sulfobus
409	6	1.9	82	2	Q5J2S9	Q5J2S9 escherichia	482	6	1.9	130	8	Q960R4	Q960R4 sulfobus
410	6	1.9	83	16	Q9PBU4	Q9PBU4 xylella fas	483	6	1.9	131	15	Q960R4	Q960R4 sulfobus
411	6	1.9	84	2	Q937P6	Q937P6 salmoneella	484	6	1.9	131	16	Q960R4	Q960R4 sulfobus
412	6	1.9	84	16	Q937P6	Q937P6 salmoneella	485	6	1.9	131	16	Q960R4	Q960R4 sulfobus
413	6	1.9	84	16	Q937P6	Q937P6 salmoneella	486	6	1.9	133	12	Q960R4	Q960R4 sulfobus
414	6	1.9	85	10	Q937P6	Q937P6 salmoneella	487	6	1.9	133	12	Q960R4	Q960R4 sulfobus
415	6	1.9	86	10	Q937P6	Q937P6 salmoneella	488	6	1.9	133	17	Q960R4	Q960R4 sulfobus
416	6	1.9	87	5	Q21640	Q21640 caenorhabd	489	6	1.9	134	12	Q960R4	Q960R4 sulfobus
417	6	1.9	87	10	Q21640	Q21640 caenorhabd	490	6	1.9	134	14	Q960R4	Q960R4 sulfobus
418	6	1.9	87	12	Q9W216	Q9W216 drosophila	491	6	1.9	134	12	Q960R4	Q960R4 sulfobus
419	6	1.9	87	12	Q9W216	Q9W216 drosophila	492	6	1.9	134	12	Q960R4	Q960R4 sulfobus
420	6	1.9	89	16	Q8XCE4	Q8XCE4 escherichia	493	6	1.9	134	16	Q960R4	Q960R4 sulfobus
421	6	1.9	89	16	Q8XCE4	Q8XCE4 escherichia	494	6	1.9	134	16	Q960R4	Q960R4 sulfobus
422	6	1.9	89	16	Q9KU77	Q9KU77 vibrio chol	495	6	1.9	135	2	Q960R4	Q960R4 sulfobus
423	6	1.9	89	16	Q9KU77	Q9KU77 vibrio chol	496	6	1.9	135	2	Q960R4	Q960R4 sulfobus
424	6	1.9	92	16	Q8X053	Q8X053 ralistonia s	497	6	1.9	135	10	Q960R4	Q960R4 sulfobus
425	6	1.9	93	2	Q8R225	Q8R225 bifidobacte	498	6	1.9	135	12	Q960R4	Q960R4 sulfobus
426	6	1.9	93	5	Q61649	Q61649 onchocerca	499	6	1.9	135	16	Q960R4	Q960R4 sulfobus
427	6	1.9	97	2	P94486	P94486 bacillus su	500	6	1.9	136	16	Q960R4	Q960R4 sulfobus
428	6	1.9	97	17	Q9HLC5	Q9HLC5 thermoplasma	501	6	1.9	137	4	Q960R4	Q960R4 sulfobus
429	6	1.9	99	10	Q943I8	Q943I8 cryza sativ	502	6	1.9	137	5	Q960R4	Q960R4 sulfobus
430	6	1.9	100	5	Q25076	Q25076 helix lucor	503	6	1.9	137	5	Q960R4	Q960R4 sulfobus
431	6	1.9	100	10	Q9SDP0	Q9SDP0 cryza sativ	504	6	1.9	138	5	Q960R4	Q960R4 sulfobus
432	6	1.9	100	12	Q84311	Q84311 human papil	505	6	1.9	138	5	Q960R4	Q960R4 sulfobus
433	6	1.9	100	16	Q9K6J3	Q9K6J3 bacillus ha	506	6	1.9	138	16	Q960R4	Q960R4 sulfobus
434	6	1.9	101	11	Q9D814	Q9D814 mus musculu	507	6	1.9	138	16	Q960R4	Q960R4 sulfobus
435	6	1.9	102	2	Q45390	Q45390 bordetella	508	6	1.9	139	17	Q960R4	Q960R4 sulfobus
436	6	1.9	103	2	Q49831	Q49831 mycobacteri	509	6	1.9	140	6	Q960R4	Q960R4 sulfobus
437	6	1.9	103	16	Q9XRT5	Q9XRT5 neisseria m	510	6	1.9	140	6	Q960R4	Q960R4 sulfobus
438	6	1.9	103	16	Q9XRT5	Q9XRT5 neisseria m	511	6	1.9	141	10	Q960R4	Q960R4 sulfobus
439	6	1.9	104	4	Q8WT08	Q8WT08 homo sapien	512	6	1.9	141	10	Q960R4	Q960R4 sulfobus
440	6	1.9	104	5	Q8WT08	Q8WT08 homo sapien	513	6	1.9	142	3	Q960R4	Q960R4 sulfobus
441	6	1.9	104	11	Q9D3I9	Q9D3I9 mus musculu	514	6	1.9	142	3	Q960R4	Q960R4 sulfobus
442	6	1.9	105	12	Q9QTY8	Q9QTY8 te virus. o	515	6	1.9	142	16	Q960R4	Q960R4 sulfobus
443	6	1.9	105	16	Q9X8C6	Q9X8C6 streptomyce	516	6	1.9	144	10	Q960R4	Q960R4 sulfobus
444	6	1.9	106	5	Q9U515	Q9U515 manduca sex	517	6	1.9	144	10	Q960R4	Q960R4 sulfobus
445	6	1.9	107	16	Q98AD2	Q98AD2 rhizobium 1	518	6	1.9	144	16	Q960R4	Q960R4 sulfobus
446	6	1.9	108	5	Q9V8Y9	Q9V8Y9 drosophila	519	6	1.9	144	16	Q960R4	Q960R4 sulfobus
447	6	1.9	109	5	Q9V8Y9	Q9V8Y9 drosophila	520	6	1.9	144	16	Q960R4	Q960R4 sulfobus
448	6	1.9	109	10	Q94CG1	Q94CG1 nicotiana t	521	6	1.9	146	6	Q960R4	Q960R4 sulfobus
449	6	1.9	109	15	Q96967	Q96967 human immun	522	6	1.9	147	3	Q960R4	Q960R4 sulfobus
450	6	1.9	109	17	Q50113	Q50113 pyrococcus	523	6	1.9	147	10	Q960R4	Q960R4 sulfobus
451	6	1.9	109	17	Q962X6	Q962X6 aeryopyrum p	524	6	1.9	148	16	Q960R4	Q960R4 sulfobus
452	6	1.9	110	5	Q962X6	Q962X6 aeryopyrum p	525	6	1.9	148	16	Q960R4	Q960R4 sulfobus
453	6	1.9	110	10	Q9ARX7	Q9ARX7 oryza sativ	526	6	1.9	148	17	Q960R4	Q960R4 sulfobus
454	6	1.9	110	10	Q9C7Y5	Q9C7Y5 arabidopsis	527	6	1.9	149	2	Q960R4	Q960R4 sulfobus

674	6	1.9	198	17	Q9YCG6	Q9YCG6 aeropyrum p	747	220	16	Q8XWB3	Q8XWB3 ralatonia s
675	6	1.9	199	17	Q27857	Q27857 methanobact	748	221	2	Q8VV17	Q8VV17 clostridium
676	6	1.9	199	17	Q8TRM3	Q8TRM3 methanosarc	749	221	5	Q97323	Q97323 plasmidium
677	6	1.9	200	16	Q9N280	Q9N280 trypanosoma	750	221	10	Q94219	Q94219 oryza sativ
678	6	1.9	200	16	Q99XK2	Q99XK2 streptococc	751	221	16	Q85344	Q85344 escherichia
679	6	1.9	200	16	Q96J77	Q96J77 rhizobium 1	752	221	10	Q851H5	Q851H5 oryza sativ
680	6	1.9	200	16	Q9ADE7	Q9ADE7 streptomyce	753	222	17	Q9H083	Q9H083 halobacteri
681	6	1.9	201	3	Q9PD3	Q9PD3 neurospora	754	222	17	Q9HKM8	Q9HKM8 thermoplasm
682	6	1.9	202	5	Q9V331	Q9V331 drosophila	755	223	5	Q9SDE0	Q9SDE0 oryza sativ
683	6	1.9	202	16	Q8XXT9	Q8XXT9 ralatonia s	756	224	5	Q46356	Q46356 caenorhabdi
684	6	1.9	203	10	Q43544	Q43544 liliun long	757	224	8	Q95896	Q95896 peromysect
685	6	1.9	203	15	Q910X4	Q910X4 human immun	758	225	16	Q8RFL4	Q8RFL4 fusobacteri
686	6	1.9	203	15	Q910X5	Q910X5 human immun	759	226	5	P91481	P91481 caenorhabdi
687	6	1.9	203	16	Q9RCX9	Q9RCX9 streptomyce	760	226	10	Q94KT3	Q94KT3 zea mays (m
688	6	1.9	204	4	Q9BR33	Q9BR33 homo sapien	761	226	10	Q9S2A3	Q9S2A3 arbidopsis
689	6	1.9	204	10	Q9CCK3	Q9CCK3 arbidopsis	762	226	16	Q9ZYR1	Q9ZYR1 rhizobium m
690	6	1.9	205	5	Q9U0H7	Q9U0H7 plasmidium	763	227	11	Q9CX65	Q9CX65 mus musculu
691	6	1.9	205	17	Q9V144	Q9V144 pyrococcus	764	228	11	Q9D2M4	Q9D2M4 mus musculu
692	6	1.9	206	17	Q9TMU2	Q9TMU2 methanosarc	765	228	16	Q9HME1	Q9HME1 pseudomonas
693	6	1.9	207	16	Q8Z1P4	Q8Z1P4 yerinia pe	766	228	17	Q25020	Q25020 archaeeoglob
694	6	1.9	207	16	Q9KX96	Q9KX96 escherichia	767	229	2	Q9EUT1	Q9EUT1 pseudomonas
695	6	1.9	209	2	Q47959	Q47959 haemophilus	768	229	12	Q9WN14	Q9WN14 mucosal dis
696	6	1.9	209	2	P74998	P74998 zymomonas m	769	229	12	Q9WN06	Q9WN06 mucosal dis
697	6	1.9	209	9	Q9T1K4	Q9T1K4 bacteriopho	770	229	12	Q9WN04	Q9WN04 mucosal dis
698	6	1.9	209	9	Q9XTM8	Q9XTM8 bacteriopho	771	229	12	Q9WN01	Q9WN01 mucosal dis
699	6	1.9	209	16	Q8XEG1	Q8XEG1 escherichia	772	229	12	Q9WN00	Q9WN00 mucosal dis
700	6	1.9	209	16	Q8YZD9	Q8YZD9 ralatonia s	773	229	12	Q9W9E2	Q9W9E2 mucosal dis
701	6	1.9	210	11	Q8VCW0	Q8VCW0 mus musculu	774	229	12	Q9W874	Q9W874 mucosal dis
702	6	1.9	210	16	Q9RKG8	Q9RKG8 streptomyce	775	229	16	Q97D24	Q97D24 clostridium
703	6	1.9	211	2	Q91BD7	Q91BD7 mycoplasma	776	230	2	Q96344	Q96344 mycobacteri
704	6	1.9	211	11	Q9CRF0	Q9CRF0 mus musculu	777	230	16	Q9K120	Q9K120 coxiella bu
705	6	1.9	211	15	Q9WAM4	Q9WAM4 feline immu	778	231	2	Q9RLV4	Q9RLV4 lactococcus
706	6	1.9	211	15	Q9WAM5	Q9WAM5 feline immu	779	231	16	Q26003	Q26003 helicobacte
707	6	1.9	211	15	Q9WAM6	Q9WAM6 feline immu	780	231	16	Q9CIN3	Q9CIN3 lactococcus
708	6	1.9	211	16	Q9PB35	Q9PB35 xyella fas	781	231	16	Q05856	Q05856 mycobacteri
709	6	1.9	212	11	Q70446	Q70446 mus musculu	782	231	16	Q9ZJF2	Q9ZJF2 helicobacte
710	6	1.9	212	17	Q8TXM7	Q8TXM7 methanopyru	783	231	16	Q9X955	Q9X955 streptomyce
711	6	1.9	213	13	Q9A42	Q9A42 populus tre	784	232	10	Q91VV7	Q91VV7 arbidopsis
712	6	1.9	213	13	Q8OX77	Q8OX77 glilichthys	785	232	16	Q8ZKA0	Q8ZKA0 salmonella
713	6	1.9	213	16	Q8XX19	Q8XX19 ralatonia s	786	233	5	Q18319	Q18319 chironomus
714	6	1.9	214	12	Q8QSV3	Q8QSV3 apple mosai	787	233	5	Q44134	Q44134 caenorhabdi
715	6	1.9	214	12	Q8QSV1	Q8QSV1 apple mosai	788	233	12	Q90V78	Q90V78 vaccinia vir
716	6	1.9	214	12	Q8QSV0	Q8QSV0 apple mosai	789	234	4	Q80375	Q80375 mycobacteri
717	6	1.9	214	12	Q8QSV0	Q8QSV0 apple mosai	790	234	4	Q961U2	Q961U2 homo sapien
718	6	1.9	214	12	Q8QSV0	Q8QSV0 apple mosai	791	234	11	Q9GEQ3	Q9GEQ3 drosophila
719	6	1.9	214	12	Q8QSV0	Q8QSV0 apple mosai	792	234	5	Q961U2	Q961U2 mus musculu
720	6	1.9	214	12	Q8QSV0	Q8QSV0 apple mosai	793	234	16	Q8YV2	Q8YV2 listeria mo
721	6	1.9	214	12	Q8QSV0	Q8QSV0 apple mosai	794	235	11	Q9XG8	Q9XG8 mus musculu
722	6	1.9	214	12	Q8QSV0	Q8QSV0 apple mosai	795	235	12	Q9D1Y4	Q9D1Y4 human coxa
723	6	1.9	214	16	Q9KFX7	Q9KFX7 bacillus ha	796	235	12	Q91L58	Q91L58 sugarcane y
724	6	1.9	214	17	Q8U4J9	Q8U4J9 pyrococcus	797	235	12	Q91L54	Q91L54 sugarcane y
725	6	1.9	215	2	Q93RL4	Q93RL4 treponema m	798	235	12	Q91L50	Q91L50 sugarcane y
726	6	1.9	215	5	Q9W4J4	Q9W4J4 drosophila	799	235	12	Q91L45	Q91L45 sugarcane y
727	6	1.9	215	6	Q28130	Q28130 bos taurus	800	235	12	Q91L41	Q91L41 sugarcane y
728	6	1.9	215	17	Q9YEV9	Q9YEV9 aeropyrum p	801	237	5	Q8T635	Q8T635 dictyosteli
729	6	1.9	216	2	Q91CR2	Q91CR2 penicillium	802	237	10	Q9AV16	Q9AV16 oryza sativ
730	6	1.9	216	16	Q8XQJ4	Q8XQJ4 ralatonia s	803	237	10	Q93YPS	Q93YPS arbidopsis
731	6	1.9	217	5	Q8S2D2	Q8S2D2 drosophila	804	237	16	Q9AD77	Q9AD77 streptomyce
732	6	1.9	217	15	Q90R19	Q90R19 human immun	805	237	17	Q9HRV1	Q9HRV1 halobacteri
733	6	1.9	218	16	Q9PRL0	Q9PRL0 xyella fas	806	238	2	Q93913	Q93913 azobacter
734	6	1.9	218	16	Q8YEX1	Q8YEX1 brucella me	807	238	5	Q9VYV6	Q9VYV6 drosophila
735	6	1.9	219	5	Q962U1	Q962U1 apodoptera	808	238	10	Q9FMX4	Q9FMX4 arbidopsis
736	6	1.9	219	10	Q9ASK8	Q9ASK8 oryza sativ	809	238	16	Q91L19	Q91L19 streptomyce
737	6	1.9	219	16	Q9RZK2	Q9RZK2 streptomyce	810	238	16	Q9RZK1	Q9RZK1 streptomyce
738	6	1.9	220	2	Q9X5H0	Q9X5H0 streptomyce	811	238	17	Q8ZM86	Q8ZM86 pyrobaculum
739	6	1.9	220	2	Q9ABE3	Q9ABE3 uncultured	812	239	12	Q9DH97	Q9DH97 rice yellow
740	6	1.9	220	2	Q9ABE2	Q9ABE2 uncultured	813	239	12	Q9DH97	Q9DH97 rice yellow
741	6	1.9	220	2	Q9ABE1	Q9ABE1 uncultured	814	239	12	Q86527	Q86527 rice yellow
742	6	1.9	220	2	Q9ABE0	Q9ABE0 uncultured	815	239	12	Q9DUR1	Q9DUR1 rice yellow
743	6	1.9	220	2	Q9AP59	Q9AP59 uncultured	816	239	12	Q9DUR0	Q9DUR0 rice yellow
744	6	1.9	220	2	Q8RT26	Q8RT26 uncultured	817	239	12	Q9DUR0	Q9DUR0 rice yellow
745	6	1.9	220	2	Q8RT16	Q8RT16 uncultured	818	239	12	Q9DUR0	Q9DUR0 rice yellow
746	6	1.9	220	5	Q9VEJ0	Q9VEJ0 drosophila	819	239	12	Q9DUR0	Q9DUR0 rice yellow

966	6	1.9	281	10	094Bt6	094Bt6 ceratostigm
967	6	1.9	281	10	022842	022842 arabidopsis
968	6	1.9	281	16	082AK7	082AK7 yerania pe
969	6	1.9	282	10	091G52	091G52 oryza sativ
970	6	1.9	283	16	09A4H3	09A4H3 caulobacter
971	6	1.9	284	10	047809	047809 enterococc
972	6	1.9	284	10	091G04	091G04 arabidopsis
973	6	1.9	284	16	082DU0	082DU0 yerania pe
974	6	1.9	284	16	0812U0	0812U0 raletonia s
975	6	1.9	284	17	0972X1	0972X1 sulfolobus
976	6	1.9	285	10	09F750	09F750 uncultured
977	6	1.9	286	11	09BMD0	09BMD0 venturia ca
978	6	1.9	286	11	09DAC6	09DAC6 mus musculu
979	6	1.9	286	16	066342	066342 clover yell
980	6	1.9	286	16	098DW5	098DW5 thizobium 1
981	6	1.9	286	16	092C23	092C23 yerania pe
982	6	1.9	286	16	09K478	09K478 streptomyce
983	6	1.9	287	5	08SRA0	08SRA0 encephalit
984	6	1.9	287	10	08JY14	08JY14 arabidopsis
985	6	1.9	287	11	08VH31	08VH31 mus musculu
986	6	1.9	288	12	066341	066341 clover yell
987	6	1.9	288	12	09DWD6	09DWD6 rat cytoomeg
988	6	1.9	288	16	091480	091480 pseudomonas
989	6	1.9	288	17	09HRF8	09HRF8 halobacteri
990	6	1.9	290	2	093CK9	093CK9 mycoplasma
991	6	1.9	290	16	08XW8	08XW8 clostridium
992	6	1.9	290	16	08UK4	08UK4 agrobacteri
993	6	1.9	291	4	096AD3	096AD3 homo sapien
994	6	1.9	291	10	024273	024273 polygonatum
995	6	1.9	291	10	08W318	08W318 vitis labru
996	6	1.9	291	16	09PQA2	09PQA2 ureaplasma
997	6	1.9	292	11	08VDB7	08VDB7 mus musculu
998	6	1.9	292	13	09DPC2	09DPC2 xenopus lae
999	6	1.9	292	16	092P88	092P88 listeria mo
1000	6	1.9	292	17	058756	058756 methanococc

ALIGNMENTS

RESULT 1
ID 09VCB2 PRELIMINARY; PRT; 968 AA.
AC 09VCB2; 01-MAY-2000 (TEMBLrel. 13, Created)
DT 01-MAY-2000 (TEMBLrel. 13, Last sequence update)
DE CG5669 Protein (L004007p).
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Ceoliker S.E., Li P.W., Hoskins R.A., Gallie R.P.,
RA Amanatides F.G., Scherer S.E., Ashburner M., Henderson S.N.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Morten U.R., Jendell M.D., Zhang Q., Chen L.X.,
RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,
RA Abiri J.F., Agbayan A., An H.-U., Andrews-Piankocch C., Baldwin D.,
RA Baillet R.M., Baou A., Baxendale U., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
RA Burks K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,

RA Foster C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Mei M.-H., Idegawa C.,
RA Jalali M., Kalush P., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasok P., Lai Y., Levitsky A.A., Li J., Li Z., Jiang Y., Lin X.,
RA Liu X., Maitel B., McIntosh T.C., McLeod M.P., McPherson D.L.,
RA Merklion G., Malhotra N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy S., Murphy L., Murry D.M., Nelson D.L.,
RA Nason S.J., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Peltman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shier B.C., Siden-Kiamos I., Simpson W., Strupski M.P., Smith T.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wastaman D.A., Weinstein G.M., Welschbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX STRAIN=BERKELEY;
RA Stapleton M., Brokstein P., Hong L., Agbayan A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA George R., Gonzalez M., Guarin H., Krommiller B., Li P., Liao G.,
RA Miranda A., Mungall C.J., Nunco J., Pacle J., Paragas V., Park S.,
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Celinker S.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: ARO03746; AAF56261.1; -
DR EMBL: AY089533; AAL90271.1; -
DR HSP: P08047; 1SP2.
DR FlyBase: FBgn0039169; CG5669.
DR InterPro: IPR000823; Znf_C2H2.
DR Pfam: PF00086; Znf_C2H2_3.
DR Pfam: PF000003; Znf_C2H2_2.
DR SMART: SM00355; Znf_C2H2_3.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 3.
DR PROSITE: PS00157; ZINC_FINGER_C2H2_2; 3.
DR DNA-binding; Metal-binding; Zinc-finger.
SQ SEQUENCE 968 AA; 102020 MW; 834CB19340C302CB CRC64;
Query Match 2.8%; Score 9; DB 5; Length 968;
Best local similarity 100.0%; Pred. No. 2.9; 0; Indels 0; Gaps 0;
Matches 9; Conservative 0; Mismatches 0;
Oy 217 ATVAGSTVT 225
Db 877 ATVAGSTVT 885
RESULT 2
ID 08S0D0 PRELIMINARY; PRT; 91 AA.
AC 08S0D0;
DT 01-JUN-2002 (TEMBLrel. 21, Created)
DT 01-JUN-2002 (TEMBLrel. 21, Last sequence update)
DE P0506X1.26 protein.
OS Oryza sativa (Japanese cultivated group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzae; Oryza.
NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;

RT "Oryza sativa nippondare(GA3) genomic DNA, chromosome 1, PAC
RT clone:PO506A10.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP003418; BAB86093.1; -
SQ SEQUENCE 91 AA; 9299 MW; 06773A6F0E18FE8 CRC64;

Query Match 2.5%; Score 8; DB 10; Length 91;
Best Local Similarity 100.0%; Pred. No. 3.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 192 TGSSGVMQ 199
DB 29 TGSSGVMQ 36

RESULT 3

ID 042049 PRELIMINARY; PRT; 120 AA.
AC 042049;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE ADP1 protein (Fragment).
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SEEDLING;
RA Desprez T., Amselem J., Chiapello H., Caboche M., Hofte H.;
RL Submitted (AUG-1993) to the EMBL/GenBank/DBJ databases.
DR EMBL; 225716; CAAB1017.1; -
DR HSRP; P24337; IHP.
DR InterPro; IPR002365; P-rich_extensin.
DR InterPro; IPR001768; Try/amy1_inhbr.
DR Pfam; PF00234; tryp_alpha_amy1; 1.
DR PRINTS; PR01217; PRICHTENGEN.
DR NON_TER 1
FT NON_TER 1
SQ SEQUENCE 120 AA; 12284 MW; BE67AECEB042CB451 CRC64;

Query Match 2.5%; Score 8; DB 10; Length 120;
Best Local Similarity 100.0%; Pred. No. 4.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 168 PTPPPPV 175
DB 18 PTPPPPV 25

RESULT 4

ID 034816 PRELIMINARY; PRT; 164 AA.
AC 034816;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE YKUD protein.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Scanlan E., Devine K.M.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;

RX MEDLINE=98044033; PubMed=9384377;

RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Boriss R., Bourcier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruch C.V., Caldwell B., Capiano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dueterhoff A., Ehrlich S.D., Emerson P.T.,
RA Dentan K.D., Errington J., Fabre C., Ferrari E., Foulger D.,
RA Fritz C., Fujita S., Fujita Y., Funa S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaeser P., Goffeau A., Golightly E.J., Grandt G.,
RA Guisepi G., Guy B.J., Haga K., Haeck J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moesl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogihara A., Oudega B., Park S.H.,
RA Paro V., Pohl T.M., Portelle D., Porwollik S., Prescott A.M.,
RA Presscan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Setor S.J., Setor P., Shin B.S., Soldo B.,
RA Sorokin A., Taccini B., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpestra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambuit R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipac A., Yamamoto H., Yamane K., Yasunoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,
RT "The complete genome sequence of the gram-positive bacterium Bacillus
RT subtilis.";
RL Nature 390:249-256 (1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ222587; CA10867.1; -
DR EMBL; Z99111; CAB13277.1; -
DR InterPro; IPR002482; Lyam.
DR Pfam; PF01476; Lyam; 1.
DR SMART; SM00257; Lyam; 1.
KW Complete proteome.

SQ SEQUENCE 164 AA; 17643 MW; 6E13A750890E7F4C CRC64;

Query Match 2.5%; Score 8; DB 16; Length 164;
Best Local Similarity 100.0%; Pred. No. 6.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 73 YOVKOGDT 80
DB 4 YOVKOGDT 11

RESULT 5

ID 09K644 PRELIMINARY; PRT; 227 AA.
AC 09K644;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Hypothetical protein BH3886.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,

RA Horikoshi K.;
 RT "Complete genome sequence of the alkaliphilic bacterium *Bacillus*
 RT *halodurans* and genomic sequence comparison with *Bacillus subtilis*.";
 RL Nucleic Acids Res. 28:4317-4331(2000).
 DR EMBL; AP001520; BAB07607.1; -.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 227 AA; 25714 MW; EEB5CFACDC3D6F3 CRC64;

Query Match 2.5%; Score 8; DB 16; Length 227;
 Best Local Similarity 100.0%; Pred. No. 8.8;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 206 ATNPVVR 213
 DB 159 ATNPVVR 166

RESULT 6
 ID 005905 PRELIMINARY; PRT; 244 AA.
 AC 005905;
 DT 01-NOV-1996 (TEMBLrel. 01, Created)
 DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
 DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
 DE Chromosome XII COSMID 8003.
 CN YLR301W OR L8003.2.
 OS *Saccharomyces cerevisiae* (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97313267; Pubmed=9169871;
 RA Johnston M., Hillier L., Riles L., Albermann K., Andre B., Ansorge W.,
 RA Benes V., Bruckner M., Delius H., Dubois E., Dusterhoft A.,
 RA Emtian K.D., Floeth M., Goffeau A., Hebling U., Heumann K.,
 RA Heuss-Neltzel D., Hilbert H., Hliger F., Kleine K., Kotter P.,
 RA Louis E.J., Messenguy F., Mewes H.W., Miosga T., Mostl D.,
 RA Muller-Auer S., Nentwich U., Obermaier B., Piravandi E., Pohl T.M.,
 RA Portetlelle D., Purnelle B., Rechmann S., Rieger M., Rinke M., Rose M.,
 RA Schaefer M., Scherens B., Scholler P., Schwager C., Schwarze S.,
 RA Underwood A.P., Urrestarazu L.A., Vandenbol M., Veraselt P.,
 RA Vierendeels F., Voet M., Volckaert G., Voss H., Wambutt R., Wedler E.,
 RA Wedler H., Zimmermann F.K., Zollner A., Hani J., Hobeisel J.D.;
 RT "The nucleotide sequence of *Saccharomyces cerevisiae* chromosome XII.";
 RL Nature 387:0-01(0).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Cherry J.W.;
 RL EMBL; U17243; AAB67346.1; -.
 DR EMBL; S0004292; YLR301W.
 SQ SEQUENCE 244 AA; 27501 MW; 80D813586A1930BB CRC64;

Query Match 2.5%; Score 8; DB 3; Length 244;
 Best Local Similarity 100.0%; Pred. No. 9.4;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 221 GSTVTSNG 228
 DB 227 GSTVTSNG 234

RESULT 7
 Q9SET1

ID Q9SET1 PRELIMINARY; PRT; 306 AA.
 AC Q9SET1;
 DT 01-MAY-2000 (TEMBLrel. 13, Created)
 DT 01-MAY-2000 (TEMBLrel. 13, Last sequence update)
 DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
 DE Cell wall-plasma membrane linker protein homolog.
 GN CMLP.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicot; Rosidae;
 OC Eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20064977; Pubmed=10598107;
 RA Geo J.H., Park A.R., Park W.J., Park O.K.;
 RT "Selection of Arabidopsis genes encoding secreted and plasma membrane
 RT proteins.";
 RL Plant Mol. Biol. 41:415-423(1999).
 DR EMBL; AF104328; AAD11796.1; -.
 DR HSSP; P24337; IHYP.
 DR InterPro; IPR003612; AAI.
 DR InterPro; IPR003882; Pistil_extensin.
 DR InterPro; IPR002965; P_rich_extensin.
 DR Pfam; PF00234; tlyp_alpha_amy1; 1.
 DR PRINTS; PR01217; PRICHEXTENSIN.
 DR SMART; SM00499; AAI; 1.
 SQ SEQUENCE 306 AA; 31609 MW; F3E6385C50F87827 CRC64;

Query Match 2.5%; Score 8; DB 10; Length 306;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 168 PTPPTPV 175
 DB 140 PTPPTPV 147

RESULT 8
 ID Q9LIE9 PRELIMINARY; PRT; 334 AA.
 AC Q9LIE9;
 DT 01-OCT-2000 (TEMBLrel. 15, Created)
 DT 01-OCT-2000 (TEMBLrel. 15, Last sequence update)
 DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
 DE Similarity to cell wall-plasma membrane linker protein.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicot; Rosidae;
 OC Eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=COLUMBIA.
 RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=COLUMBIA.
 RX MEDLINE=20363099; Pubmed=10907853;
 RA Nakamura Y.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
 RT Sequence features of the regions of 4,251,695 bp covered by ninety PL,
 RT TAC and BAC clones.";
 RL DNA Res. 7:217-221(2000).
 DR EMBL; AP001306; BAB03061.1; -.
 DR HSSP; P24337; IHYP.
 DR InterPro; IPR003612; AAI.
 DR InterPro; IPR003882; Pistil_extensin.
 DR InterPro; IPR002965; P_rich_extensin.
 DR InterPro; IPR001768; Tlyp_alpha_amy1_inhbr.

DR Pfam; PF00234; try_alpha_aml; 1.
 DR PRINTS; PR01217; PRICEXTENSIN.
 DR PRINTS; PR01218; PSTEXTENSIN.
 DR SMART; SM00499; AAI; 1.
 SQ SEQUENCE 334 AA; 34669 MW; F99159A17AD70524 CRC64;

Query Match 2.5%; Score 8; DB 10; Length 334;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 168 PTPPPVV 175
 DB 168 PTPPPVV 175

RESULT 9
 ID Q9353 PRELIMINARY; PRT; 376 AA.
 AC Q9353;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Cell wall-plasma membrane linker protein.
 GN
 OS Brassica napus (Rape).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Brassica.
 NCBI_TaxID=3708;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COBRA;
 RA Goodwin W.G., Pallas J.A., Jenkins G.I.
 RT "Transcripts of a gene encoding a putative cell wall-plasma membrane
 linker protein are specifically cold-induced in Brassica napus."
 RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL; X94976; CAA64425.1; -
 DR HSSP; P24337; IHYP.
 DR InterPro; IPR003612; AAI.
 DR InterPro; IPR001768; Try/amy1_inhbr.
 DR Pfam; PF00234; try_alpha_aml; 1.
 DR SMART; SM00499; AAI; 1.
 SQ SEQUENCE 376 AA; 38667 MW; 7E12AD838938A4F8 CRC64;

Query Match 2.5%; Score 8; DB 10; Length 376;
 Best Local Similarity 100.0%; Pred. No. 14;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 168 PTPPPVV 175
 DB 175 PTPPPVV 182

RESULT 10
 ID Q9UG05 PRELIMINARY; PRT; 88 AA.
 AC Q9UG05;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE BK989H1.1 (Novel protein) (Fragment).
 GN BK989H1.1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NCBI_TaxID=9606;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC Ramsey H.;
 RA Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; Z83851; CAB62962.1; -
 FT NON_TER 1 1
 SQ SEQUENCE 88 AA; 9699 MW; C2ED8FE7AB67CFA1 CRC64;

Query Match 2.2%; Score 7; DB 4; Length 88;
 Best Local Similarity 100.0%; Pred. No. 43;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 184 PVTAPF 190
 DB 57 PVTAPF 63

RESULT 11
 ID Q9CDX3 PRELIMINARY; PRT; 101 AA.
 AC Q9CDX3;
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE 50S ribosomal protein L24.
 GN RPLX OR L2408.
 OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
 OC Streptococcaceae; Lactococcus.
 NCBI_TaxID=1360;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=IL1403;
 RX MEDLINE=21235186; PubMed=11337471;
 RA Bolotin A., Wincker P., Mauer S., Jallion O., Malarne K.,
 RA Weisenbach J., Ehrlich S.D., Sorokin A.;
 RT "The complete genome sequence of the lactic acid bacterium Lactococcus
 lactis ssp. lactis IL1403."
 RL Genome Res. 11:731-753 (2001).
 DR EMBL; AE006437; AAK06186.1; -
 DR InterPro; IPR000302; KOW_motif.
 DR InterPro; IPR003256; Ribosomal_L24.
 DR Pfam; PF00467; KOW; 1.
 DR Prodom; PD001677; Ribosomal_L24; 1.
 DR TIGRPFAM; TIGR01079; rplX_bact; 1.
 DR PROSITE; PS01108; RIBOSOMAL_L24; UNKNOWN_1.
 DR Ribosomal protein, Complete proteome.
 KW
 SQ SEQUENCE 101 AA; 10877 MW; B4CAF87BE72C9021 CRC64;

Query Match 2.2%; Score 7; DB 16; Length 101;
 Best Local Similarity 100.0%; Pred. No. 48;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 279 VKTGDV 285
 DB 3 VKTGDV 9

RESULT 12
 ID Q8S142 PRELIMINARY; PRT; 111 AA.
 AC Q8S142;
 DT 01-JUN-2002 (TREMBlrel. 21, Created)
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE P0042A10.23 protein.
 GN P0042A10.23.
 OS Oryza sativa (japonica cultivar-group).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzaceae; Oryza.
 NCBI_TaxID=39947;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. NIPPONBARE;
 RA Sasaki T., Matsunoto T., Yamamoto K.;
 RT "Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1, PAC
 clone: P0042A10."
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AP003343; BAB90081.1; -

```

SQ SEQUENCE 111 AA; 11245 MW; 212977985AB0D90A CRC64;
Query Match 2.2%; Score 7; DB 10; Length 111;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 178 PAPVAP 184
DB 44 PAPVAP 50

RESULT 13
Q8RUE2 PRELIMINARY; PRT; 122 AA.
ID Q8RUE2;
AC Q8RUE2;
DT 01-JUN-2002 (TReMBLrel. 21, Created)
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
DE P0703B11.17 protein (P0485B12.9 protein).
GN P0703B11.17 OR P0485B12.9.
OS Oryza sativa (Japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponebare (GA3) genomic DNA, chromosome 1, PAC
clone: P0703B11.1";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa (Japonica cultivar-group) genomic DNA, chromosome 1, PAC
clone: P0485B12.9";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP003302; BAB85298.1;
DR EMBL; AP003348; BAB86479.1;
SQ SEQUENCE 122 AA; 13319 MW; D988C108BA8FA070 CRC64;

Query Match 2.2%; Score 7; DB 10; Length 122;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 220 AGSTVTS 226
DB 23 AGSTVTS 29

RESULT 14
Q9ZWT6 PRELIMINARY; PRT; 133 AA.
ID Q9ZWT6;
AC Q9ZWT6;
DT 01-MAY-1999 (TReMBLrel. 10, Created)
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
DE 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
GN ZMGRI1 protein.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PAC clade;
OC Panicoidae; Andropogoneae; Zea.
OX NCBI_TaxID=4577;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LEAF SHEATH;
RA Ogawa M., Kusano T., Koizumi N., Katsumi M., Sano H.;
RT "Gibberellin-responsive genes: high level of transcript accumulation
in leaf sheath meristematic tissue from Zea mays L.";
RN [2]
RP SEQUENCE FROM N.A.

QY 168 PTPPPV 174
DB 26 PTPPPV 32

RESULT 15
Q9ST25 PRELIMINARY; PRT; 133 AA.
ID Q9ST25;
AC Q9ST25;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DE 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
GN ZMGRI1 protein.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PAC clade;
OC Panicoidae; Andropogoneae; Zea.
OX NCBI_TaxID=4577;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LEAF SHEATH;
RA Ogawa M., Kusano T., Koizumi N., Katsumi M., Sano H.;
RT "Gibberellin-responsive genes: high level of transcript accumulation
in leaf sheath meristematic tissue from Zea mays L.";
DR EMBL; AB018588; BAA74804.1;
DR HSSP; P24337; IHP.
DR InterPro; IPR003612; AAI.
DR InterPro; IPR001768; Try/amy1_inhbr.
DR Pfam; PF00234; try_alpha_amy1; 1.
DR SMART; SM00499; AAI; 1.
SQ SEQUENCE 133 AA; 13465 MW; F2AF4E1584FF9631 CRC64;

Query Match 2.2%; Score 7; DB 10; Length 133;
Best Local Similarity 100.0%; Pred. No. 62;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 168 PTPPPV 174
DB 26 PTPPPV 32

RESULT 16
Q9FWP4 PRELIMINARY; PRT; 136 AA.
ID Q9FWP4;
AC Q9FWP4;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DE 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
GN OSUNBA0015015.11.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.

```

RC STRAIN=CV. NIPPONBARE;
 RA Buell C.R., Yuan Q., Moffat K.S., Hill J.N., Jenkins C.N., Burr P.C.,
 RA Hsiao J., Zismann V., Pai G., Bowman C.L., Fujii C.Y., Vanaken S.E.,
 RA Bowman C.L., Craven B., Utterback T.R., Khalak H., Feldblyum T.V.,
 RA Quackenbush J., White O., Salzberg S.L., Fraser C.M.;
 RT "Oryza sativa chromosome 10 BAC OSUNBA0015015 genomic sequence."
 RT Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 RL EMBL; AC026758; AAG13475.1; -
 DR HSSP; P24337; IHP.
 DR InterPro; IPR003612; AAI.
 DR InterPro; IPR001768; Try/amy1_inhbr.
 DR Pfam; PF00234; tryp_alpha_amy1; 1.
 DR SMART; SM00499; AAI; 1.
 SQ SEQUENCE 136 AA; 13982 MW; 9F785B01C3115NAF CRC64;

Query Match 2.2%; Score 7; DB 10; Length 136;
 Best Local Similarity 100.0%; Pred. No. 63;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 158 PTPPVVV 164
 DB 37 PTPPVVV 43

RESULT 17
 ID P91181 PRELIMINARY; PRT; 141 AA.

AC P91181;
 DT 01-MAY-1997 (TREMBlrel. 03, Created)
 DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE Hypothetical 15.0 kDa protein.
 GN C50F2.7.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Felodermidae; Caenorhabditis.
 OC NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RX MEDLINE=99069613; PubMed=9851916;
 RA None;
 RT "Genome sequence of the nematode C. elegans: a platform for
 RT investigating biology. The C. elegans Sequencing Consortium.";
 RL Science 282:2012-2018 (1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 DU 2., Le T.T.;
 RT "The sequence of C. elegans cosmid C50F2.";
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RA "Direct Submission."
 RT Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
 RL EMBL; U80445; AAB37800.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 141 AA; 15050 MW; 1F655AC7BAE14E99 CRC64;

Query Match 2.2%; Score 7; DB 5; Length 141;
 Best Local Similarity 100.0%; Pred. No. 66;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 222 STVTSNG 228
 DB 58 STVTSNG 64

RESULT 18
 Q9FWP0 PRELIMINARY; PRT; 142 AA.

AC Q9FWP0;
 DT 01-MAR-2001 (TREMBlrel. 16, Created)
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Putative lipid transfer protein.
 GN OSUNBA0015015.15.
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Eriarthoidae; Oryzaceae; Oryza.
 OC NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. NIPPONBARE;
 RA Buell C.R., Yuan Q., Moffat K.S., Hill J.N., Jenkins C.N., Burr P.C.,
 RA Hsiao J., Zismann V., Pai G., Bowman C.L., Fujii C.Y., Vanaken S.E.,
 RA Bowman C.L., Craven B., Utterback T.R., Khalak H., Feldblyum T.V.,
 RA Quackenbush J., White O., Salzberg S.L., Fraser C.M.;
 RT "Oryza sativa chromosome 10 BAC OSUNBA0015015 genomic sequence."
 RT Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 RL EMBL; AC026758; AAG13488.1; -
 DR HSSP; P24337; IHP.
 DR InterPro; IPR003612; AAI.
 DR InterPro; IPR001768; Try/amy1_inhbr.
 DR Pfam; PF00234; tryp_alpha_amy1; 1.
 DR SMART; SM00499; AAI; 1.
 SQ SEQUENCE 142 AA; 14393 MW; C3E8AA4F34A63AD1 CRC64;

Query Match 2.2%; Score 7; DB 10; Length 142;
 Best Local Similarity 100.0%; Pred. No. 66;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 158 PTPPVVV 164
 DB 37 PTPPVVV 43

RESULT 19
 ID Q8THE9 PRELIMINARY; PRT; 165 AA.

AC Q8THE9;
 DT 01-JUN-2002 (TREMBlrel. 21, Created)
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Multiple resistance/pH regulation related protein E (Na⁺/H⁺
 DE antiporter).
 GN MKPE OR M44568.
 OS Methanosarcina acetivorans.
 OC Archaea; Euryarchaeota; Methanococci; Methanosarcinales;
 OC Methanosarcinaceae; Methanosarcina.
 OC NCBI_TaxID=2214;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C2A / ATCC 35395 / DSM 2834;
 RX MEDLINE=21929760; PubMed=11932238;
 RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., Macdonald P.,
 RA Fitzhugh W., Calvo S., Engels R., Smirnov S., Ancoor D., Brown A.,
 RA Allen N., Naylor J., Strange-Thomann N., DeAtellano K., Johnson R.,
 RA Linton L., McEwan P., McKernan K., Talamas J., Tittell A., Ye W.,
 RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,
 RA Hedderich R., Ingram-Smith C., Kuetner H.C., Krzycki J.A.,
 RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
 RA Springer T.A., Umayan L.A., White O., White R.H., de Macario E.C.,
 RA Ferry J.G., Jarelli K.F., Jing H., Macario A.J.L., Paulsen I.,
 RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
 RA Metcalf W.W., Birren B.;
 RT "The genome of Methanosarcina acetivorans reveals extensive metabolic
 RT and physiological diversity.";
 RL Genome Res. 12:532-542 (2002).
 DR EMBL; AF011178; AAM07907.1; -
 KW Complete proteome.
 SQ SEQUENCE 165 AA; 18327 MW; DCAB72532F74E05E CRC64;

Query Match 2.2%; Score 7; DB 17; Length 165;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 LGLIFGV 24
DB 29 LGLIFGV 35

RESULT 20
0926C3 PRELIMINARY; PRT; 176 AA.
AC 0926C3; 01-DEC-2001 (TEMBLrel. 19, Created)
DT 01-DEC-2001 (TEMBLrel. 19, Last sequence update)
DE 01-MAR-2002 (TEMBLrel. 20, Last annotation update)
GN Probable peptidoglycan-associated lipoprotein precursor.
OS PAL OR R02738 OR SMC02942.
OC Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=11396507; PubMed=11481430;
RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batur J.,
Boisard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
Godrie T., Goffeau A., Kahn D., Kiss E., Lelaire V., Maury D.,
Pohl T., Portetelle D., Puehler A., Purnelle B., Rampeger U.,
Renard C., Thebault P., Vandenbol M., Weidner S., Gilbert F.,
RT "Analysis of the chromosome sequence of the legume symbiont
RT Sinorhizobium meliloti strain 1021.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
DR EMBL: A591791; CA047317.1;
DR InterPro: IPR001145; Bac_OmpA.
DR Pfam: PF00651; OmpA, 1.
DR ProDom: PD000930; Bac_OmpA; 1.
KM SEQUENCE 176 AA; 18735 MW; A6F162CB35042268 CRC64;
SQ Complete proteome.

Query Match 2.2%; Score 7; DB 16; Length 176;
Best Local Similarity 100.0%; Pred. No. 80;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 30 LAGCASK 36
DB 30 LAGCASK 36

RESULT 21
09XW11 PRELIMINARY; PRT; 177 AA.
AC 09XW11; 01-NOV-1999 (TEMBLrel. 12, Created)
DT 01-NOV-1999 (TEMBLrel. 12, Last sequence update)
DE 01-DEC-2001 (TEMBLrel. 19, Last annotation update)
GN V54E2A.4 protein.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Rhabditinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Lloyd C.R.;
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RT none;
RT "Genome sequence of the nematode C. elegans: A platform for
RT investigating biology.";

RL Science 282:2012-2018(1998).
DR EMBL: AL032646; CNA21686.1;
SQ SEQUENCE 177 AA; 20062 MW; E3BA47B50CCFA329 CRC64;

Query Match 2.2%; Score 7; DB 5; Length 177;
Best Local Similarity 100.0%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 STSGSGS 47
DB 62 STSGSGS 68

RESULT 22
099YU1 PRELIMINARY; PRT; 179 AA.
AC 099YU1; 01-JUN-2001 (TEMBLrel. 17, Created)
DT 01-JUN-2001 (TEMBLrel. 17, Last sequence update)
DE 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
GN Hypothetical protein SPY1538.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SF370 / ATCC 700294 / SEROTYPE M1;
RX MEDLINE=21192684; PubMed=11296296;
RA Fieretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
Pimenta X., Szate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
DR EMBL: AB006586; AA34332.1;
DR InterPro: IPR003398; Cons_hypoth95.
DR InterPro: IPR002052; N6_Mcase.
DR Pfam: PF03602; Cons_hypoth95; 1.
DR TIGRfam: TIGR00095; Cons_hypoth95; 1.
DR PROSITE: PS00092; N6_MTASE; UNKNOWN; 1.
KM Hypothetical protein; Complete proteome.
SQ SEQUENCE 179 AA; 19567 MW; BD3A72BFF541137 CRC64;

Query Match 2.2%; Score 7; DB 16; Length 179;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 52 GSGGLAI 58
DB 51 GSGGLAI 57

RESULT 23
09SKR4 PRELIMINARY; PRT; 180 AA.
AC 09SKR4; 01-DEC-2001 (TEMBLrel. 19, Created)
DT 01-DEC-2001 (TEMBLrel. 19, Last sequence update)
DE 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
GN ART5 protein (Fragment).
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RA Kuehl M., Glowacki G., Haag F., Koch-Noite F.;
RT "Molecular cloning and characterization of human ART5, a secretory
RT ecto-mono (ADP-ribosyl) transferase from testis.";
RT Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL: AJ294468; CAC07429.1; -
DR InterPro: IPR000768; ART_family.
DR Pfam: PF01129; ART; 1.

FT NON_TER 1 1
FT NON_TER 180 180
SQ SEQUENCE 180 AA; 20101 MW; 8DC89A96B4B9452 CRC64;

Query Match 2.2%; Score 7; DB 6; Length 180;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 212 RRFGRAT 218
Db 151 RRFGRAT 157

RESULT 24

082FF8 PRELIMINARY; PRT; 189 AA.

AC 082FF8; 01-MAR-2002 (TREMBLrel. 20, Created)

DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)

DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

GN Putative membrane protein.

OS Yersinia pestis.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

NCBI_Taxid=63;

RA Yersinia

RA Yersinia

RA Yersinia

RA Yersinia

RA Yersinia

RA Yersinia

RA Yersinia

RA Yersinia

RA Yersinia

RA Yersinia

RA Yersinia

RA Yersinia

RA Yersinia

RA Yersinia

RA Yersinia

RA Yersinia

RA Yersinia

RA Yersinia

RA Yersinia

RA Yersinia

RA Yersinia

RA Bourbon H.M., Allemand E., Soret J., Tazi J.;
RT "Characterization of the Drosophila ortholog of the mammalian pre-mRNA
RT splicing factor SC35";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.

RP SEQUENCE FROM N.A. (LONG AND SHORT ISOFORMS).

RC STRAIN=BERKELEY;

RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amaral L.D., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Milos J.L.G.,

RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfankuch C., Baldwin D.,

RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.Y., Benos P.V., Berman B.P., Bhattacharya D., Bolshakov S.,

RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotier P.,

RA Burris K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,

RA Fostler C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,

RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Hostin N.L., Harvey D., Heilmann T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Mei M.-H., Ibegwam C.,

RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merkulov G., Milehina N.V., Mobarry C., Morris J., Moirer A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusslein D.R., Pacle J.M.,

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Smith H.,

RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smet T.,

RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

RA Switkes R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,

RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,

RA Williams S.M., Woodedge T., Worley K.C., Wu D., Yang S., Yao Q.A.,

RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

RA Gibbs X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,

RA Gibbs R.A., Myers B.W., Rubin G.M., Venter J.C.,

RT "The genome sequence of Drosophila melanogaster";

RL Science 287:2185-2195 (2000).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=BERKELEY;

RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,

RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,

RA Gonzalez M., Guatin H., Li P., Liao G., Miranda A., Mungall C.J.,

RA Nuno J., Pacle J., Paragas V., Park S., Phouanavong S., Wan K.,

RA Yu C., Lewis S.E., Rubin G.M., Celniker S.,

RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.

CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A LONG FORM (SHOWN HERE) AND A

CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.

DR EMBL: AF232775; AAF43415.1; -

DR EMBL: AB003636; AAF51192.1; -

DR EMBL: AB003636; AAF51193.1; -

DR EMBL: AY069584; AAL39729.1; -

DR HSSP: P11940; 1CVJ.

DR Flybase: FBgn0040286; SC35.

DR InterPro: IPR000504; RNA_rec_mot.

DR Pfam: PF00076; rrm; 1.

DR SMART: SM00360; RRM; 1.

DR PROSITE: PS0102; RRM; 1.

DR PROSITE: PS00030; RRM_RNP_1; UNKNOWN_1.

RT Alternative splicing

FT VARSPLIC 1 83 MISSING (IN SHORT ISOFORM).

SQ SEQUENCE 195 AA; 21400 MW; 65E922F8C34C7536 CRC64;

Query Match 2.2%; Score 7; DB 5; Length 195;

Best Local Similarity 100.0%; Pred. No. 88;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 51 SSGCGLA 57
Db 179 SSGCGLA 185

RESULT 26

OBXXR6 PRELIMINARY; PRT; 205 AA.
AC 08XXR6;
DT 01-MAR-2002 (T-EMBLrel. 20, Created)
DT 01-MAR-2002 (T-EMBLrel. 21, Last sequence update)
DE Hypothetical protein RSC2047.
GN RSC2047 OR RS03612.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GM11000;
RA MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Arlet M., Billault A., Brotier P., Camus J.C., Catroico L.,
RA Chandler M., Choiane N., Claudel-Renard C., Cumac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,
RA Weisenbach J., Boucher C.A., Whalen M., Wincker P., Levy M.,
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RT Nature 415:497-502 (2002).
DR EMBL: AL646068; CDD:5754.1;
DR InterPro: IPR002667; IPR_1isomerase.
DR InterPro: IPR000086; NUDIX_hydrolase.
DR Pfam: PF00293; NUDIX.1.
DR PRINTS: PR00502; NUDIXFAMILY.
DR Prodom: PD004109; IPR_1isomerase; 1.
KM Hypothetical protein; Complete proteome.
SQ SEQUENCE 205 AA; 22905 MW; 011998A4555B1B10 CRC64;

Query Match 2.2%; Score 7; DB 16; Length 205;
Best Local Similarity 100.0%; Pred. No. 93;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 27

OB1130 PRELIMINARY; PRT; 208 AA.
AC 081130;
DT 01-NOV-1998 (T-EMBLrel. 08, Created)
DT 01-NOV-1998 (T-EMBLrel. 08, Last sequence update)
DT 01-MAR-2002 (T-EMBLrel. 20, Last annotation update)
DE CCAT-box binding factor HAP3 Homolog.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC Eustoids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WS-0;
RA MEDLINE=98319234; PubMed=9657152;
RA Loezan T., Ohto K.M., West M.A., Lo R., Kwong R.W.,
RA Yanagishi K., Fischer R.L., Goldberg R.B., Harada J.J.;
RT "Arabidopsis LEAFY COTYLEDON1 is sufficient to induce embryo
development in vegetative cells.";
RL Cell 93:1195-1205 (1998).

DR EMBL: AF036684; AAC39488.1; --
DR HSSP: P19267; 1A7W.
DR InterPro: IPR003958; CBFA_NFYB_domain.
DR InterPro: IPR003957; CBFA_NFYB_topis.
DR InterPro: IPR004822; Histone_core.
DR Pfam: PF00808; CBFD_NFYB_HMF.1.
DR PRINTS: PR00615; COATSUBUNTA.
SQ SEQUENCE 208 AA; 22693 MW; 4E2D249AE325DDB CRC64;

Query Match 2.2%; Score 7; DB 10; Length 208;
Best Local Similarity 100.0%; Pred. No. 94;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 314 YVDPPLTV 320
Db 105 YVDPPLTV 111

RESULT 28

09SPD8 PRELIMINARY; PRT; 208 AA.
AC 09SPD8;
DT 01-MAY-2000 (T-EMBLrel. 13, Created)
DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)
DT 01-MAR-2002 (T-EMBLrel. 20, Last annotation update)
DE T26P17.20
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC Eustoids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C., Khan S.,
RA Kim C., Altafi H., Bei Q., Chin C., Chlou J., Choi E., Conn L.,
RA Conway A., Gonzalez A., Hansen N., Howing B., Koo T., Lam B., Lee J.,
RA Lenz C., Li J., Liu A., Liu J., Liu S., Mukharasy N., Nguyen M.,
RA Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaveri A.,
RA Toriumi M., Vaysberg M., Yu G., Federpiel N.A., Theologis A.,
RA Ecker J.R.;
RT "Genomic sequence for Arabidopsis thaliana BAC T26P17 from chromosome
I.";
RT Submitted (NOV-1999) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Ecker J.R.;
RN Submitted (NOV-1999) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C., Khan S.,
RA Kim C., Altafi H., Bei Q., Chin C., Chlou J., Choi E., Conn L.,
RA Conway A., Gonzalez A., Hansen N., Howing B., Koo T., Lam B., Lee J.,
RA Lenz C., Li J., Liu A., Liu J., Liu S., Mukharasy N., Nguyen M.,
RA Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaveri A.,
RA Toriumi M., Vaysberg M., Yu G., Davis R., Federpiel N., Theologis A.,
RA Ecker J.;
RN Submitted (DEC-1999) to the EMBL/Genbank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Cheuk R., Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,
RA Khan S., Kim C., Altafi H., Bei Q., Chin C., Chlou J., Choi E., Conn L.,
RA Conn L., Conway A., Gonzalez A., Hansen N., Howing B., Koo T., Lam B.,
RA Lee J., Lenz C., Li J., Liu A., Liu J., Liu S., Mukharasy N.,
RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,
RA Thaveri A., Toriumi M., Vaysberg M., Yu G., Davis R., Federpiel N.,
RA Theologis A., Ecker J.;
RN Submitted (JUN-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL: AC013482; AAF16537.1; --
DR HSSP: P19267; 1A7W.
DR InterPro: IPR003958; CBFA_NFYB_domain.
DR InterPro: IPR003957; CBFA_NFYB_topis.
DR InterPro: IPR004822; Histone_core.
DR Pfam: PF00808; CBFD_NFYB_HMF.1.

DR PRINTS, PRO0615; CCAATSUBUNTA.
SQ SEQUENCE 208 AA; 22679 MW; 0DEB469D9F8BEFDC CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 94;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 314 YVDPLTV 320
Db 105 YVDPLTV 111

RESULT 29
Q9S2U3 PRELIMINARY; PRT; 211 AA.
AC Q9S2U3;
DT 01-MAY-2000 (TEMBLrel. 13, Created)
DT 01-MAY-2000 (TEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
DE Putative membrane protein.
GN SC02042 OR SC466.11C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomycetes.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Frazer A., Goble A., Hidalgo J., Hornaby T., Howarth S.,
RA Huang C.-H., Kleser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitz B., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
coelicolor A3(2)."
RL Nature 417:141-147(2002).
DR EMBL; AL096884; CAB51434.1; -
SQ SEQUENCE 211 AA; 21223 MW; 4A8123837C5683B8 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 95;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 215 GATVAG 221
Db 50 GATVAG 56

RESULT 30
Q9FYQ4 PRELIMINARY; PRT; 212 AA.
AC Q9FYQ4;
DT 01-MAR-2001 (TEMBLrel. 16, Created)
DT 01-MAR-2001 (TEMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)
DE Hypothetical protein (P0011G08.31 protein).
GN P0011G08.31.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
clone: P0433F09.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.

RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
clone: P0011G08.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP002539; BAB08181.1; -
DR EMBL; AP003225; BAB64659.1; -
KW Hypothetical protein.
SQ SEQUENCE 212 AA; 23424 MW; C5B5EB20D459DA24 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 95;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 43 SSGSGSHR 49
Db 160 SSGSGSHR 166

RESULT 31
Q9VT29 PRELIMINARY; PRT; 213 AA.
AC Q9VT29;
DT 01-MAY-2000 (TEMBLrel. 13, Created)
DT 01-MAR-2001 (TEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
DE CG14414 protein.
GN CG14414.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachyera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.B., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-Y., Andrews-Pfannkoch C., Baldwin D.,
RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Eickler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegam C.,
RA Jalali M., Kalish F., Karpen G.H., Ke Z., Kemison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Matrei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mileshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Mutzy D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheefer F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 225 AA; 25873 MW; 27A28FB3CCE2188A CRC64;

Query Match 2.2%; Score 7; DB 17; Length 225;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 118 GDLKVR 124
Db 123 GDLKVR 129

RESULT 34

Q92KU4 PRELIMINARY; PRT; 226 AA.
AC 092KU4:
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DE Hypothetical protein R00142.
GN R00142 OR SMC04110.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN=1021.
RX MEDLINE=21396507; PubMed=11481430;
RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
RA Boistard P., Becker A., Boutry M., Cadieu B., Dreano S., Gloux S.,
RA Godre T., Goffeau A., Kahn D., Kise E., Lelaure V., Maury D.,
RA Pohl T., Potteville P., Puehler A., Purnelle B., Ransperger U.,
RA Renard C., Thebaud P., Vandenbol M., Weidner S., Galibert F.;
RT "Analysis of the chromosome sequence of the legume symbiont
RT Sinorhizobium meliloti strain 1021.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
DR EMBL; AL591782; CAC41529.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 226 AA; 24278 MW; CF261B891D7B48E CRC64;

Query Match 2.2%; Score 7; DB 16; Length 226;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 30 LAGCASK 36
Db 20 LAGCASK 26

RESULT 35

Q9H5P8 PRELIMINARY; PRT; 228 AA.
AC 09H5P8:
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE CDNA: FLJ23189.1; c10ne LNC12061.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.

RC TISSUE=LUNG;
RA Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T., Matsumura K.,
RA Nakajima Y., Mizuno T., Morinaga M., Tanigami A., Fujiwara T., Ono T.,
RA Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y., Ota T., Suzuki Y.,
RA Iwagishi M., Nishii T., Shibahara T., Tanaka T., Nakamura Y.,
RA Togeiri T., Sugano S.,
RT "NEO human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK026842; BAB15572.1; -.

SQ SEQUENCE 228 AA; 26292 MW; 171D0DCD1364E01F CRC64;

Query Match 2.2%; Score 7; DB 4; Length 228;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 143 QSSRPV 149
Db 166 QSSRPV 172

RESULT 36

Q9KLV7 PRELIMINARY; PRT; 241 AA.
AC 09KLV7:
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DE Hypothetical protein VCA0634.
GN VCA0634.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN=EL TOR N16961 / SEROTYPE O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey B.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Esmolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Uitterlind T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae.";
RL Nature 406:477-483(2000).
DR EMBL; AE004393; AAF96535.1; -.
DR TIGR; VCA0634; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 241 AA; 27259 MW; 1AE3F947FCD04B27 CRC64;

Query Match 2.2%; Score 7; DB 16; Length 241;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 82 SKIAQRY 88
Db 192 SKIAQRY 198

RESULT 37

Q8TUT2 PRELIMINARY; PRT; 241 AA.
AC 08TUT2:
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Cobalamain-5-phosphate synthase.
GN COBS OR MK1671.
OS Methanopyrus kandleri.
OC Archaea; Euryarchaeota; Methanopyri; Methanopyrales; Methanopyraceae;
OC Methanopyrus.
OX NCBI_TaxID=2320;
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN=AV19 / DSM 6324 / JCM 9639;
RX MEDLINE=21927647; PubMed=11930014;
RA Slesarev A.I., Mezheva K.V., Makarova K.S., Polunin N.N.,
RA Shcherbinina O.V., Shakhova V.V., Belova G.I., Astarov L.,
RA Natale D.A., Rogozin I.B., Tatunov R.L., Wolf Y.I., Stetter K.O.,
RA Malyn A.G., Koonin E.V., Kozayavkin S.A.;
RT "The complete genome of hyperthermophile Methanopyrus kandleri AV19

RT and monophyly of archaeal methanogens.";
 Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649(2002).
 DR EMBL; AB010456; AM02884.1; --
 KW Complete proteome.
 SQ SEQUENCE 241 AA; 25264 MW; E1C25302FB7FEB6 CRC64;

Query Match 2.2%; Score 7; DB 17; Length 241;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 54 GGLAIGS 60
 DB 109 GGLAIGS 115

RESULT 38

O9HNY3 PRELIMINARY; PRT; 249 AA.
 AC O9HNY3;
 DT 01-MAR-2001 (TEMBLrel. 16, Created)
 DT 01-MAR-2001 (TEMBLrel. 16, Last sequence update)
 DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
 DE ABC transport protein.
 GN TRP2 OR UNG1893G.
 OS Halobacterium sp. (strain NRC-1).
 OC Archaea; Euryarchaeota; Halobacteriales;
 OC Halobacteriaceae; Halobacterium.
 OX NCBI_TaxId=64091;

RN
 RP SEQUENCE FROM N.A.
 RX MEDLINE=2050448; PubMed=11016950;
 RA NG W.V., Kennedy S.P., Mahaitas G.G., Bergquist B., Pan M.,
 RA Shukla H.D., Laeky S.R., Baliga N.S., Thoreson V., Sirogma J.,
 RA Swartzell S., Weir D., Hall J., Dahl T.A., Weir T.A., Goo Y.A.,
 RA Leitauer B., Keller K., Cruz R., Danson M.J., Hough D.W.,
 RA Maddocks D.G., Jablonksi P.E., Krebs M.P., Angewine C.M., Dale H.,
 RA Aiem B., Fretas T., Hou S., Daniels C.J., Spudich J.L., Jung K.-H.,
 RA Ebhardt H., Lowe T.M., Liang P., Dennis P.P., Omer A.D.,
 RA "Genome sequence of Halobacterium species NRC-1";
 RA Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
 RL EMBL; AE005088; AAC20087.1; --
 DR InterPro; IPR003593; AAA_ATPase.
 DR InterPro; IPR003439; ABC_transport.
 DR Pfam; PF00005; ABC_tran; 1.
 DR ProDom; PD000006; ABC_transport; 1.
 DR SMART; SM00382; AAA; 1.
 KW Complete proteome.
 SQ SEQUENCE 249 AA; 26679 MW; 06F96B1A8D2AE19 CRC64;

Query Match 2.2%; Score 7; DB 17; Length 249;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 215 GTATVAG 221
 DB 60 GTATVAG 66

RESULT 39

O51373 PRELIMINARY; PRT; 259 AA.
 AC O51373;
 DT 01-JUN-1998 (TEMBLrel. 06, Created)
 DT 01-JUN-1998 (TEMBLrel. 06, Last sequence update)
 DT 01-MAR-2002 (TEMBLrel. 20, Last annotation update)
 DE Hypothetical protein BB0412.
 GN BB0412.
 OS Borrelia burgdorferi (Lyme disease spirochete).
 OS Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
 OX NCBI_TaxId=139;
 RN
 RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 35210 / B31;
 MEDLINE=98065943; PubMed=9403685;
 RX Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
 RA Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,
 RA Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,
 RA Peterson J., Karpavavage A.R., Quackenbush J., Salzberg S., Hanson M.,
 RA van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,
 RA Urebedts S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
 RA Smith H.O., Venter C.;
 RT "Genomic sequence of a Lyme disease spirochete, Borrelia
 burgdorferi";
 RL Nature 390:580-586(1997).
 DR EMBL; AB01146; AAC66791.1; --
 DR TIGR; BB0412; --
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 259 AA; 30809 MW; 2137010F9140DOFF CRC64;

Query Match 2.2%; Score 7; DB 16; Length 259;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 99 NNINSSY 105
 DB 109 NNINSSY 115

RESULT 40

O94238 PRELIMINARY; PRT; 269 AA.
 AC O94238;
 DT 01-FEB-1997 (TEMBLrel. 02, Created)
 DT 01-FEB-1997 (TEMBLrel. 02, Last sequence update)
 DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)
 DE Hypothetical 30.2 kDa protein.
 GN F55A4.4.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Rhabditinae; Caenorhabditis.
 OX NCBI_TaxId=6239;
 RN
 RP SEQUENCE FROM N.A.
 RX STRAIN=BRISTOL N2; PubMed=9851916;
 RA None;
 RT "Genome sequence of the nematode C. elegans: a platform for
 RT investigating biology. The C. elegans Sequencing Consortium.";
 RL Science 282:2012-2018(1998).

RT
 RL
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RA Bradshaw H.;
 RT "The sequence of C. elegans cosmid F55A4.";
 RT Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RA Waterston R.;
 RT "Direct Submission";
 RL EMBL; U67949; AB07562.1; --
 DR HSSP; Q91836; 1D12.
 DR InterPro; IPR001159; DS_RBD.
 DR Pfam; PF00035; dsm; 2.
 DR SMART; PF000358; DSRM; 2.
 DR PROSITE; PS0137; DS_RBD; 2.
 KW Hypothetical protein.
 SQ SEQUENCE 269 AA; 30169 MW; 2028FE3245563191 CRC64;

Query Match 2.2%; Score 7; DB 5; Length 269;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	315	VDP	TVL	321
Db	14	VDP	TVL	20

Search completed: July 8, 2003, 11:20:55
Job time : 111 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_p2n model

Run on: July 3, 2003, 13:47:57 ; Search time 2083 Seconds
(without alignments)
4498.849 Million cell updates/sec

Title: US-10-018-706-2
Perfect score: 1670
Sequence: 1 MTWTAINSQKPKIRLGL.....LFEFRISNGVYVDLTVLK 322

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters: -DB=tbl
-O=/cgn2.1/USPTO_SPOOL/US10018706/runat_30062003_091104_23794/app_query.fasta_1.519
-DB=genemb1 -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -STRAT=1 -END=1 -MATRIX=blomsun62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=40 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MITLEN=0 -MAXLEN=2000000000
-USER=US10018706@cgn.1.1.1687@runat_30062003_091104_23794 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEUSER -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmb1.*
1: gb_ba.*
2: gb_hcg.*
3: gb_in.*
4: gb_cm.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_scs.*
12: gb_sy.*
13: gb_un.*
14: gb_vl.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_scs.*
28: em_un.*

29: em_vl.*
30: em_hcg_hum.*
31: em_hcg_inv.*
32: em_hcg_other.*
33: em_hcg_mus.*
34: em_hcg_pin.*
35: em_hcg_rtd.*
36: em_hcg_mam.*
37: em_hcg_vrt.*
38: em_sy.*
39: em_hgo_hum.*
40: em_hgo_mus.*
41: em_hgo_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1670	100.0	969	6 AX063563	AX063563 Sequence
2	1670	100.0	48328	6 AX067452	AX067452 Sequence
3	1666	99.8	966	6 AX063565	AX063565 Sequence
4	298	17.8	2910	1 PSEBPOS	D26134 Pseudomonas
5	298	17.8	13031	1 AB004782	AB004782 Pseudomonas
6	289	17.3	58996	1 AB034704	AB034704 Ruditiviva
7	286	17.1	3122	1 AB006073	AB006073 Pseudomonas
8	282	16.9	5876	1 AF421351	AF421351 Azotobact
9	253	15.1	10026	1 AB003925	AB003925 Xylella f
10	249.5	14.9	12956	1 AB006198	AB006198 Pasteurel
11	245	14.7	3539	1 PPY19122	Y19122 Pseudomonas
12	238	14.3	9720	1 AB000369	AB000369 Escherich
13	238	14.3	55175	1 EC028375	U28375 Escherichia
14	238	14.3	26668	1 AP002563	AP002563 Escherich
15	234.5	14.0	10773	1 AB012272	AB012272 Xanthomon
16	234	14.0	10822	1 AB004905	AB004905 Pseudomon
17	232.5	13.9	22400	1 AB008839	AB008839 Salmonell
18	232.5	13.9	230050	1 AL627277	AL627277 Salmonell
19	231.5	13.9	1552	1 AF260132	AF260132 Pseudomon
20	227.5	13.6	2609	1 AF244357	AF244357 Coxidia
21	225	13.5	9360	1 AB005516	AB005516 Escherich
22	222.5	13.3	207050	1 AL646063	AL646063 Ralstonia
23	219.5	13.1	11435	1 U32753	U32753 Haemophilus
24	219.5	13.1	11555	1 AB011804	AB011804 Xanthomon
25	219.5	13.1	14193	1 AB002061	AB002061 Deinococc
26	218.5	13.1	10592	1 AB013686	AB013686 Yersinia
27	218.5	13.1	220050	1 AJ414156	AJ414156 Yersinia
28	213.5	12.8	1433	1 EC00CMTLPD	L07669 Escherichia
29	213.5	12.8	10573	1 AB005502	AB005502 Escherich
30	213.5	12.8	11457	1 AB000358	AB000358 Escherich
31	213.5	12.8	72221	1 EC029579	U29579 Escherichia
32	213.5	12.6	270365	1 AP002562	AP002562 Escherich
33	210	12.6	274050	1 AL627276	AL627276 Salmonell
34	209	12.5	1290	1 SDU6131	AJ006131 Salmonell
35	209	12.5	2179	6 RR210638	Sequence
36	209	12.5	20984	1 AB008833	AB008833 Salmonell
37	208.5	12.5	10104	1 AB004139	AB004139 Vibrio ch
38	206	12.3	1612	1 EC0RPOS	D17549 Escherichia
39	205	12.3	12362	1 AB009547	AB009547 Brucella
40	203	12.2	329861	1 NMA552491	ALI62756 Neisseria
41	203	12.2	349619	1 AP002966	AP002966 Mesorhizo
42	198	11.9	11085	1 AB002497	AB002497 Neisseria
43	198	11.9	349980	6 AX044032	Sequence
44	198	11.9	349980	6 AX044033	Sequence
45	197.5	11.8	3892	12 EC0FV1AB	Z21706 E. coli/C.fr

RESULT 1

ALIGNMENTS

AX063563
 LOCUS AX063563 969 bp DNA linear PAT 24-JAN-2001
 DEFINITION Sequence 1 from Patent WO0100838.
 ACCESSION AX063563
 VERSION AX063563.1 GI:12541291
 KEYWORDS
 SOURCE Moraxella catarrhalis.
 ORGANISM Moraxella catarrhalis.
 Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;
 Moraxella.

REFERENCE
 1 (bases 1 to 969)
 AUTHORS Thomaz, J.S.
 TITLES Cloning of BapB10 antigen from moraxella (brahameilla) catarrhalis
 JOURNAL Patent: WO 0100838-A 1 04-JAN-2001;
 SMITHKLINE BEECHAM BIOLOGICALS (S.A.)
 FEATURES
 source 1..969
 Location/Qualifiers
 /organism="Moraxella catarrhalis"
 /db_xref="taxon:480"

BASE COUNT 265 a 214 c 242 g 248 t

ORIGIN

Alignment Scores:
 Pred. No.: 1.07e-90 Length: 969
 Score: 1670.00 Matches: 322
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: Gaps: 0

US-10-018-706-2 (1-322) x AX063563 (1-969)

QY 1 MetThValThrIleAlaIleAsnSerGlnAsnGlnLysProIleLysArgLeuGlyLeu 20
 DB 1 ATATACGTGACGATGACATCAATTCACAAATCCAAACCCATCAAGCATTCGGCTTG 60

QY 21 IlePhgGlyValIleThrThrCysIleLeuAlaGlyCysAlaSerLysProThrTyrAsn 40
 DB 61 ATTTTGGTGTATGATCACACCTTGATTTTGCGAGATGCGAGTGAAGCCAACTTAAT 120

QY 41 SerThrSerGlySerGlySerHisArgThrSerGlySerGlyLysLeuAlaIleGlySer 60
 DB 121 AGTACCTCAGGTTCCGGGACATCATCTGACTTCAGTTCAGTGGTTGGCAATAGTTCA 180

QY 61 GlnValIleThrAspSerGlnGlyValProAsnArgTyrGlnValLysGlnGlyAspThr 80
 DB 181 CAGGTTATACGACGACAGTACGAGCGGTACAAATCGCTATCAGGTGAAGCAGCGGATACT 240

QY 81 ValSerLysIleAlaGlnArgTyrGlyLeuAsnTyrArgLuiIleGlyHisIleAsnAsn 100
 DB 241 GTCAGTATGATTCCTCAGCGTTATGATTAATGGCGTGAAGATGGACATTAAATAT 300

QY 101 LeuAsnSerSerTyrThrIleTyrThrGlyGlnTyrPheuThrPheuTyrAspLeu 120
 DB 301 CTAAATATGACGATATACGATTATACAGTCATAGGTGACTTATGTCAGGATCTC 360

QY 121 LysValArgGlnArgSerIleSerSerGlyValAsnThrLysThrProSerProVal 140
 DB 361 AAGTCGTAGAGCTATCATCAGCTCTGAGTGTATACAGTCAACACCTTCGCTGTG 420

QY 141 AlaValGlnSerSerArgProProValGlnGlnHisProAlaValGlnLysProThrPro 160
 DB 421 GCGGTTCAGTACGACGACACCAACGATACAGATCTTCGCGTACAAACCCACGCCA 480

QY 161 ProValValValValLysLysProThrProThrProProValValGlnGlnProAlaPro 180
 DB 481 CCGTGTGTGTGTAAACCAACCCACACGACCTCCGCTGTGTGTCGACGACGCCA 540

QY 181 ValAlaProProValThrGlnAlaProPheAlaThrGlySerSerGlyValIleGlnPhe 200
 DB 541 GTTGCCCAACGATGACAGAAAGCACATTTGCGGAGTACAGGCGGTGAATGCAATTT 600

QY 201 ArgTyrProValGlyAlaThrAsnProValValArgArgPheGlyThrAlaThrValAla 220

DB 601 CGGTATCTGTGTGTGGACCAATCCAGTGTTCAGCGCTTGTGGTACGGGACAGTGGCC 660

QY 221 GlySerThrValThrSerAsnGlyMetTyrPheSerGlyValArgAspGlyLeuIleAsn 240
 DB 661 GCGTCACTGTTCACGATATGCAATGCTGTGTTCGTGAGAGATGGCCATTTAATTAAC 720

QY 241 AlaSerAsnAlaGlyThrValIleGlnAlaAspHisAsnMetArgValAsnIleVal 260
 DB 721 GCCAGTATGACGACGACGATTCATTCAGCTGATACCAATATGACGCGGACGATTTGTG 780

QY 261 IleGlnHisThrAsnGlyPheValSerSerTyrIleHisIleLysAspAlaGlnValLys 280
 DB 781 ATTCAGATACCAATGATGATTTGTTTCAGCTATATATCAATTAAGACGCTCAAGTTAAA 840

QY 281 ThrGlyAspThrValArgThrGlyGlnArgIleAlaSerMetLysAsnGlnProSerGly 300
 DB 841 ACAGGCGATACGGTGGTACCGGTACCGCTATTCGAAACATGAAATCAGCCAAAGCGGT 900

QY 301 AlaAlaLeuPheGlnPheArgIleSerArgAsnGlyValTyrValAspProLeuThrVal 320
 DB 901 GCGGCACTATTGAATTTAGATTCTTCAAGAAATGCGGTATGTGATTCATTGACAGTA 960

QY 321 LeuLys 322
 DB 961 CTTAAA 966

RESULT 2
 AX067452 48326 bp DNA linear PAT 24-JAN-2001
 LOCUS AX067452
 DEFINITION Sequence 27 from Patent WO0078968.
 ACCESSION AX067452
 VERSION AX067452.1 GI:12545072
 KEYWORDS
 SOURCE Moraxella catarrhalis.
 ORGANISM Moraxella catarrhalis.
 Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;
 Moraxella.

REFERENCE
 1 (bases 1 to 48328)
 AUTHORS Lagace, R.E., Patterson, C. and Berg, K.L.
 TITLES Nucleotide sequences of moraxella catarrhalis genome
 JOURNAL Patent: WO 0078968-A 27 28-DEC-2000;
 Incyte Genomics, Inc. (US)
 FEATURES
 source 1..48328
 Location/Qualifiers
 /organism="Moraxella catarrhalis"
 /db_xref="taxon:480"

BASE COUNT 14211 a 9322 c 10802 g 13992 t

ORIGIN

Alignment Scores:
 Pred. No.: 8.02e-89 Length: 48328
 Score: 1670.00 Matches: 322
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: Gaps: 0

US-10-018-706-2 (1-322) x AX067452 (1-48328)

QY 1 MetThValThrIleAlaIleAsnSerGlnAsnGlnLysProIleLysArgLeuGlyLeu 20
 DB 40850 ATATACGTGACGATGACATCAATTCACAAATCCAAACCCATCAAGCATTCGGCTTG 40909

QY 21 IlePhgGlyValIleThrThrCysIleLeuAlaGlyCysAlaSerLysProThrTyrAsn 40
 DB 40910 ATTTTGGTGTGTATGATCACACCTTGATTTTGCGAGATGCGAGTGAAGCCAACTTAAT 40969

QY 41 SerThrSerGlySerGlySerHisArgThrSerGlySerGlyLysLeuAlaIleGlySer 60
 DB 40970 AGTACCTCAGGTTCCGGGACATCATCTGACTTCAGTTCAGTGGTTGGCAATAGTTCA 41029

QY 61 GlnValIleThrAspSerGlnGlyValProAsnArgTyrGlnValIleGlyGlnGlyAspThr 80
DB 41030 CAGGTTATACCGAGACAGTCAAGGCGTACCAATCGCTATAGGTGAAGCAGGCGCATCT 41089
QY 81 ValSerIleIleAlaGlnArgTyrGlyLeuAsnTyrArgIleGlyHisIleAsnAsn 100
DB 41090 GTCAGTAAGATTGCTCAGCGTTATGATTAATGCGTGAAGTGAACACATTAATTAAT 41149
QY 101 LeuAsnSerSerTyrThrIleTyrThrGlnGlnTyrLeuThrLeuTyrPsrSerGlyAspLeu 120
DB 41150 CTAAATAGAGATTATGATTATTAACAGGCAATGCTGCTTATAGTCAAGTGAATCTC 41209
QY 121 LysValArgGluArgSerIleSerSerGlyValAsnThrAlaHisThrProSerProVal 140
DB 41210 AAGGTCCGAGCGTATGATCACTGCTGCTGATTAATACAGCTACACACCTTGCTG 41269
QY 141 AlaValGlnSerSerArgProProValGlnGlnHisProAlaValGlnLysProThrPro 160
DB 41270 GCGGTTCAGTCAAGCAGACCAAGACAGACATCTCCGTACAAAACCCAGCCCA 41329
QY 161 ProValValValValLysLysProThrProThrProProValValGlnGlnProAlaPro 180
DB 41330 CCGTGTGTGTGTAATAAAACCCACACGACCTCCCTGTGTGTCAGCAGCAGCAGCA 41389
QY 181 ValAlaProProValThrGluAlaProPheAlaThrGlySerSerGlyValMetGlnPhe 200
DB 41390 GTTGCCTCCACAGTACAGACAGCAATCTCCAGGCTGCTCAGGCGGAGATGCAATTT 41449
QY 201 ArgTyrProValGlyValAlaThrAsnProValValArgArgPheGlyThrAlaThrValAla 220
DB 41450 CCGTATCTGTGTGTGTCAGCAATCGATCGATTCGACGCTTGTGTGTAAGCGGACAGTGGCC 41509
QY 221 GlySerThrValThrSerAsnGlyMetTyrPheSerGlyValArgAspGlyAspLeuIleAsn 240
DB 41510 GGCTCAACGTTACCGATATGCGATGCGATGCTTTCGACGAGATGCGCATTAATTAAC 41569
QY 241 AlaSerAsnAlaGlyThrValIleGlnAlaAspHisAsnMetAspGlyAlaSerIleVal 260
DB 41570 GCCAGTAAGCAGGACGACATCTTCAAGCTGATCAATATGACCGGCGGAGATATTGTG 41629
QY 261 IleGlnHisThrAsnGlyPheValSerSerTyrIleHisIleLysAspAlaGlnValLys 280
DB 41630 ATTCAAGATACCAATGATTTGTTTCAAGCTATATCAATTAAGACGCTCAAGTTAA 41689
QY 281 ThrGlyAspThrValArgThrGlyGlnArgIleAlaSerMetLysAsnGlnProSerGly 300
DB 41690 ACAGGCGATACGCTGCTGACCGCTGACCGATTCAGACATGAAAAATCGCCAGCGGT 41749
QY 301 AlaAlaLeuPheGluPheArgIleSerArgAsnGlyValTyrValAspProLeuThrVal 320
DB 41750 GCGGCACTATTGTAATTGAAATTTCTAGAAATGGCGTATGATTCATTAACAGTA 41809
QY 321 LeuLys 322
DB 41810 CTTAAA 41815

RESULT 3
AX063565
LOCUS AX063565 966 bp DNA linear PAT 24-JAN-2001
DEFINITION Sequence 3 from Patent WO0100838.
ACCESSION AX063565
VERSION AX063565.1 GI:12541292
KEYWORDS
SOURCE
ORGANISM
Moraxella catarrhalis.
Moraxella catarrhalis
Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;
Moraxella.
REFERENCE
1 (bases 1 to 966)
AUTHORS
TITLE
JOURNAL
Thomard,J.S.
Cloning of babbl10 antigen from moraxella (branchamella) catarrhalis
Patent: WO 0100838-A 3 04-JAN-2001;
SMITHKLINE BEECHAM BIOLOGICALS (S.A.)

FEATURES
source Location/Qualifiers
1..966
/organism="Moraxella catarrhalis"
/db_xref="taxon:480"
BASE COUNT 263 a 214 c 242 g 247 t
ORIGIN
Alignment Scores:
Pred. No.: 1,85e-90 Length: 966
Score: 1666.00 Matches: 321
Percent Similarity: 99.69% Conservative: 0
Best Local Similarity: 99.69% Mismatches: 1
Query Match: 99.76% Indels: 0
DB: Gaps: 0
US-10-018-706-2 (1-322) x AX063565 (1-966)
QY 1 MetThrValThrIleAlaIleAsnSerGlnAsnGlnLysProIleLysArgLeuGlyLeu 20
DB 1 ATGACTGTGACGATACCAATCAATTCACAAATCAAAACCCATCAAGCATTTGGCTTG 60
QY 21 IlePheGlyValIleThrThrCysIleLeuAlaGlyCysAlaSerLysProThrTyrAsn 40
DB 61 ATTTTGTGTGATCAACACTTGCATTTTGGCAGATGTGCGATAGCCATTAAT 120
QY 41 SerThrSerGlySerGlySerHisArgThrSerGlySerGlyGlyLeuAlaIleGlySer 60
DB 121 AGTACCTCAGGTTCCGAGCATCTGCTACTTCAGGTTCAGGTGTTTGGCAATAGTTCA 180
QY 61 GlnValIleThrAspSerGlnGlyValProAsnArgTyrGlnValIleGlyGlnGlyAspThr 80
DB 181 CAGGTTATACCGAGACAGTCAAGGCGTACCAATCGCTATCAGGTGAAGCAGGCGCATCT 240
QY 81 ValSerIleIleAlaGlnArgTyrGlyLeuAsnTyrArgIleGlyHisIleAsnAsn 100
DB 241 GTCAGTAAGATTGCTCAGCGTTATGATTAATTAATGCGTGAAGTGAACACATTAATTAAT 300
QY 101 LeuAsnSerSerTyrThrIleTyrThrGlnGlnTyrLeuThrLeuTyrPsrSerGlyAspLeu 120
DB 301 CTAAATAGCGGTTATCGATTTTATACAGTCAATGCTGCTGCTTATAGTCAAGTATCTC 360
QY 121 LysValArgGluArgSerIleSerSerGlyValAsnThrAlaHisThrProSerProVal 140
DB 361 AAGGTCCGAGCGTATGATCACTGCTGCTGATTAATACGCTACACACCTTGCGCTGTG 420
QY 141 AlaValGlnSerSerArgProProValGlnGlnHisProAlaValGlnLysProThrPro 160
DB 421 GCGGTTCAGTCAAGCAGACCAAGACAGACATCTGCTGCTGCTGCTGCTGCTGCTGCTG 480
QY 161 ProValValValValLysLysProThrProThrProProValValGlnGlnProAlaPro 180
DB 481 CCGTGTGTGTGTAATAAAACCCACACGACCTCCCTGTGTGTTGACGACGACCA 540
QY 181 ValAlaProProValThrGluAlaProPheAlaThrGlySerSerGlyValMetGlnPhe 200
DB 541 GTTGCCCAACAGTACAGAGACCAATTTGCCAGGCTGCTCAGGCGGATGCAATTT 600
QY 201 ArgTyrProValGlyValAlaThrAsnProValValArgArgPheGlyThrAlaThrValAla 220
DB 601 CGCTATCTGTGTGTGTCAGCAATTCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 660
QY 221 GlySerThrValThrSerAsnGlyMetTyrPheSerGlyValArgAspGlyAspLeuIleAsn 240
DB 661 GGCTCACTGTTTACGATATGCGATGTGTTTTTTCGACGAGATGCGCATTAATTAATTAAC 720
QY 241 AlaSerAsnAlaGlyThrValIleGlnAlaAspHisAsnMetAspGlyAlaSerIleVal 260
DB 721 GCCAGTAAGCAGGACAGTCAATTCAGCTGATCAATATGACCGGCGGAGATATTGTG 780
QY 261 IleGlnHisThrAsnGlyPheValSerSerTyrIleHisIleLysAspAlaGlnValLys 280
DB 781 ATTCAGATACCAATGATTTGTTTCACTATATCAATTAATTAAGACGCTCAAGTTAA 840


```

Db      891 GGGTGGCATGGCTCCGAGCGGTACT-----CTGATCGCGCGTTT----- 932
      219 VALAAGlySerThrValThrSerAnGlyMetTrpPheSerGlyThrAspGlyAspLeu 238
      933 ---GCTCAAGCGAAGTTGAATTAAGGATTGATATAGCCCGCTAATTGGCCAGCCT 969
Qy      239 ILeAsnAlaSerAsnAlaGlyThrValIleGlnAlaAspHisAsnMetAspGlyAlaSer 258
      990 GTCTGCGTGGCGCTGCTGGTGGACCGTTGATACCGCGGTAGTGGTTGGCGGGCTACGGC 1049
Db      259 -----IleValIleGlnHisThrAnGlyPheValSerSerTrpIleHisIleLeuAsp 276
      1050 GAGTTGTCATCATCTCAACACACAGACCTACGTGATGCTTACCGCTCACACCCGACAG 1109
Qy      277 AlAGlnVallyThrGlyAspThrValArgThrGlyGlnArgIleAlaSerMetLeuAsn 296
      1110 CTGCTGTCGGGGAGAGGCAACAGCTCAAGTAGGGCAATTCGATGCCAGATGGGCTCTC 1169
Qy      297 GlnProSerGlyAlaAlaLeuPheGluPheArgIleSerArgAnGlyValTyValAsp 316
      1170 ACAGGACCGCATCGGGTGAAGCTGCACCTTCGAGATTCCGCCAGGTAAGCCTGTCGAT 1229
Qy      317 ProLeuThrValLeu 321
      1230 CCACTGCATATTATTG 1244
Db
Qy
Db
RESULT 5
AE004782/c 13031 bp DNA linear BCT 30-AUG-2000
LOCUS pseudomonas aeruginosa PA01, section 343 of 529 of the complete
DEFINITION genome.
ACCESSION AE004782 AE0044091
VERSION AE004782.1 GI:9949772
KEYWORDS
SOURCE
ORGANISM Pseudomonas aeruginosa.
Pseudomonas aeruginosa.
Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
Pseudomonas.
REFERENCE 1 (bases 1 to 13031)
AUTHORS Stover,C.K., Pham,X.Q., Erwin,A.L., Mizoguchi,S.D., Warrenner,P.,
Hickey,M.J., Brinkman,F.S.L., Hufnagle,W.O., Kowalik,D.J., Lagrou,M.,
Garber,R.L., Goltry,L., Tolentino,E., Westbrook-Wadman,S., Yuan,Y.,
Brody,L.L., Coulter,S.N., Polger,K.R., Kas,A., Larbig,K., Lim,R.,
Smith,K., Spencer,D., Wong,G.K., Wu,Z. and Paulsen,I.T.
Complete genome sequence of Pseudomonas aeruginosa PA01, an
opportunistic pathogen
Nature 406 (6799), 959-964 (2000)
JOURNAL
MEDLINE 20437337
PUBMED 10984043
REFERENCE 2 (bases 1 to 13031)
AUTHORS Stover,C.K., Pham,X.Q.T., Erwin,A.L., Mizoguchi,S.D., Warrenner,P.,
Hickey,M.J., Brinkman,F.S.L., Hufnagle,W.O., Kowalik,D.J.,
Lagrou,M., Garber,R.L., Goltry,L., Tolentino,E.,
Westbrook-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N.,
Polger,K.R., Kas,A., Larbig,K., Lim,R.M., Smith,K.A., Spencer,D.H.,
Wong,G.K.-S., Wu,Z., Paulsen,I.T., Reizer,J., Sater,M.H.,
Hancock,R.E.W., Lory,S. and Olson,M.V.
Direct Submission
Submitted (16-MAY-2000) Department of Medicine and Genetics,
University of Washington Genome Center, University of Washington,
Box 352145, Seattle, WA 98195, USA
TITLE
JOURNAL Location/Qualifiers
FEATURES
source 1..13031
/organism="Pseudomonas aeruginosa"
/strain="PA01"
/db_xref="taxon:287"
72..1142
/gene="PA3615"
72..1142
/gene="PA3615"
/codon_start=1
/transl_table=11
gene
CDS
/feature="hypoetical protein"
/protein_id="AAG07003.1"
/db_xref="GI:9949773"
/transl_table="MPEADYLSRHFOTSGTDIAKVDLEALAAAPGSPINATLYRE
MPTTTPMAQADRNRRDAKIMQTLREMHASVLEQFRKRTVTFGSAKPTVEHP
VALARLGEELARLYDMLVTTGAGGIGMAAHRGAGLENSLGFNITLPEQOHANHTVDG
SGNLSHFHFFPLRKLFEVKEADALVLCPGGFTLDEALBDVLTIVQKGPVLVIVLD
OGGRVHEALFEFMOLENNHYILPADRLKRLVHSADAVEILQFYRNPFSSPML
KQTPVRLNHALNEALHAHFASLCSGGGQQQAYSEQDEDEFPNLTFLAAVE
NRRDQRLREILDYINLPENNMD"
complement (1164..1625)
/gene="PA3616"
complement (1164..1625)
/gene="PA3616"
complement (1164..1625)
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAG07004.1"
/db_xref="GI:9949774"
/transl_table="MAIVDTPVAVRARMADLLARREHRAELSRLLRORGASAEILD
PALDRLAERGLDESRYLESFTASRRSGHPURFEEIAROGSLPADIERALGACEV
DMSAQLRBYWRKRPALPDARERKQGFRLAYRGISMSISRLNGRBD"
complement (1631..2671)
/gene="PA3617"
complement (1631..2671)
/gene="PA3617"
complement (1631..2671)
/codon_start=1
/transl_table=11
/product="Reca protein"
/protein_id="AAG07005.1"
/db_xref="GI:9949775"
/transl_table="MDENKRALAAALGOIEROGFGAVRWGDHEROALPATSGSL
GIDIALGTGLPKGRIVEIYGPBSSGKTLITLSVIAERKQATCAFPDAERADBDY
AKLGVNVDLIVSQDTEQALLETDMVRSNANAVITIVSVAALVPAEELGEGND
AHVGLQARLMSQALRKITGNIKANCLVFINQIRKIKIEMVNEPPTTGGNALKEFYA
SVRLDIRRTGAVEGDEVSETRVAVNKKVSPPRQAEFQILYKGIYRTGEIIDL
GVOLGIVKSGAMVYQSGKIGQKXNNAKYLEDNBEISVLEKTIIRDQLAKSGPVK
ADAERVADEAD"
complement (2805..3311)
/gene="PA3618"
complement (2805..3311)
/gene="PA3618"
complement (2805..3311)
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAG07006.1"
/db_xref="GI:9949776"
/transl_table="MTLMSDLTALAARGLSELRYLGEQVSTAESCCTGGGIAEATRI
PGSGWFEAGVTVYSNRKTLQLEVESLFPVAVGAVSREVEAMVGAORHSGARFAY
AYSGVAGPDGSGPEKRVGTWMLAMADGERLVSERQFPNDRDRAVRRQYATALTGLLR
MSNGENPG"
3459..4466
/gene="PA3619"
3459..4466
/gene="PA3619"
complement (3459..4466)
/gene="PA3619"
complement (3459..4466)
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAG07007.1"
/db_xref="GI:9949777"
/transl_table="NNIAPLKYSLSLIALSLSPVASIPMAHACTANILAPASSLPAL
GGRLVHYSMEYEGDSSNLYLHDFKSKRLOINOGKWNLEDMMNAPFSRDPGRVLTGMC
RONGAHVYPMALIGCTQAPSNITTAIGENNEPKPSFDRQVYLKHEGDIRLATLVFN
GDGSYGVSAWKAIVTDGWSSTESMPFLTPSGKVVVYATAGSLRVRNRLNNGQVCA
LATPAAGRDYVYVVDYAYPLSRTOPAGNDQLAMVNSPGRPALPLNHCQDCA
SDAAVVEDYLIIFSSTSPFTYSLILGIDIAARVRLDPAQINLPDGRQKLGASYTAA
R"
4592..7159
/gene="mutS"
4592..7159
/gene="mutS"
complement (4592..7159)
/gene="mutS"
complement (4592..7159)

```


[illegible]

```

/gene="ORF277"
/notes="similar to ubiquinone methyltransferase UblE"
/citation={3}
/codon_start=1
/translation={1}
/protein_id="BAA94025.1"
/db_xref="GI:7416776"
/translation="MRAAGCPYKVSQSDPTLIDRIPPMNPPTISRSQVSYRE
RERORIRRPYSVAVARYDMLMDLMSGIRLKRIRYRMAAQAQGHVLDAGTDS
VALMAADRRVYVYVDSAEAMAVGQKGAHVDQVGSASQPLADASVDITLISG
IRNATRIDVALREIRHYLKTGRTLCLESTPAWMLRPFNLSTFTYIPRGMATIAS
EAATLVESIRKRDORGFAMTISAQFESVRWHDLSFGIACVHGTAAATPACT
A"
/gene=".6499
/gene="ORF468"
/notes="hypothetical enzyme for heme synthesis
(coproporphyrinogen III)"
/citation={3}
/codon_start=1
/translation={1}
/protein_id="BAA94026.1"
/db_xref="GI:7416777"
/translation="MRTPTDHPDHPDPSVYLKTIANLDAGASMCWQSPASPLVQ
GMLGNSPBARLYEALRGADRPDEPLVASRPPCAHCFCCDDIDRAQDEYI
DVLVAGVEIRVAVAGVLRGLLOPHLGGTASRLTGSLARLVQAMQSNOMRDEY
IDMSCEDPRVVSOTMEVLRGLGFRVSLVDLDVQOALIGRQSAALIDVCEI
ARASGIEYINLDMIGLPROTASMCATVRLDAMGERIRLARVYRHPHAPVQVAI
DTSLEPSPALMAELARASAEMLNAGRWITGADQFVLEHDLSPRMDGRSLISY
TAAPIYPMIGLGVAGVEIDGAMFWMNDSQAAMNLRHLPSQARPATPTPSQOR
AVERRLCTLELAADAVGLEDGRLAREAGLVAVLDRIYVTEAGHNLHALCT
DIUEPQAPRGSRARCC"
/gene="7230
/gene="ORF247"
/notes="ORF247"
/citation={3}
/codon_start=1
/translation={1}
/product="hypothetical cytochrome with 4 heme binding
sites"
/protein_id="BAA94027.1"
/db_xref="GI:7416778"
/translation="MRLNWRATATASFAAVGLAFLALPALASDADAASAPAPV
APGAAPVAPASEAPTHRLPLDNLPOAPNPQEPILFGYLDPAFESVVRKHLGLY
PCSQCHVLPUNKQPRALVAAPVAVSLPHRGRMMCLDCHVARDMLRTLDGRVDF
DOSQVCGCHGPRHRDWAFAHGKRAANNQGEROIYACTHCHDHPKIPREPAFP
PVRAGLERPALTPAHAAQWORAASGAR"
/gene=".8195
/gene="ORF321"
/notes="ORF321"
/citation={3}
/codon_start=1
/translation={1}
/product="hypothetical electron transfer protein with 4
Fes centers"
/protein_id="BAA94028.1"
/db_xref="GI:7416779"
/translation="MSRQPSDPPDDPGQPMTPRRPFLACVATAGAVTAAATAM
GGOTLDAMGSLQDHQKQRTKDEIHAAALRIENNRRTGVDRCELDLPRTVFG
YANISCKCKYRCQVCKENNLGRDSQITRVLQMPGTDLERHAYTDKQVP
VPMHTFQLPVQCCNDPVCYACGCTKATMEPDITVVDVDMCIGCTACTCTGAKR
HEMTFQLPVQCCNDPVCYACGCTKATMEPDITVVDVDMCIGCTACTCTGAKR
VFENLDPKSEIKMVLNNVFLKEDLGEKFWYPSDDRVSPA"
/gene=".9418
/gene="ORF408"
/notes="ORF408"
/citation={3}
/codon_start=1

```

```

Alignment Scores:
Pred. No.: 5,4e-07 Length: 58996
Score: 289.00 Matches: 94
Percent Similarity: 43.20% Conservative: 49
Best Local Similarity: 28.40% Mismatches: 143
Query Match: 17.31% Indels: 48
DB: 1 Gaps: 9

US-10-018-706-2 (1-322) x AB034704 (1-58996)
QY 10 GlnAsnGlnLysProIleLeuValLeu-----GlyLeuIlePheGlyValIle 25
Db 53379 GAGAACTCCCGTCCCGACCCGACCTTCCGCGGCGCTTCCCTGCTGATGAGCGGTG 53438
QY 26 ThrThrCysIleLeuAlaGlyCysAlaSerLysProThrTyraSerThrSergLys 45
Db 53439 GCCCGCGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 53495
QY 46 GlySerHisArgThrSergLysGlyGlyLeuAlaIleGlySergValIle----- 63
Db 53496 CGCAGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 53555
QY 64 -----ThrAspSergGlnValProAsnArgThrGlnValLysGlnGlyAsp 79
Db 53556 CCGCGCGCGCGCGCAGAAACCTGGGCAAGCCCGGCTACTACCGCGTCAAGCCCGG 53615
QY 80 ThrValSerLysIleAlaGlnArgTyrglyLeuAsnTrpArgGlnIleGlyHisIleAsn 99
Db 53616 ACCGTATCCGATCCGATCCGATCCGATCCGATCCGATCCGATCCGATCCGATCC 53675
QY 100 AsnLeuAsnSergLysThrIleTyrglyGlyGlyLeuThrLeuThrSergLysAsp 119
Db 53676 AACCTCAGAAACCCGACCGCATCGAAGTGGCCAGTCTGCGCGCTGCGCGCGG 53734
QY 120 LeuLysValArgLysGlySergLysSergLysValAsnThrAlaHisThrProSergPro 139
Db 53735 TCCGACCGCCAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 53793
QY 140 ValAlaValGlnSergLysArgProValGlnGlnHisProAlaValGlnLysProThr 159
Db 53794 GTCGG-----CGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 53835
QY 160 ProProValValValLysLysProThrProThrProProValValGlnGlnProAla 179
Db 53836 -----ProValThrGlnLysProProAlaThr 192
QY 180 ProValAlaPro-----ProValThrGlnLysProProAlaThr 192
Db 53866 CCGCAGCGCGCAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 53924
QY 193 GlySerSergLysValMetGlnPheArgTyrglyProValGlyValAlaThrAsnProValAlaArg 212
Db 53925 GACGCGCAGCAGCAGCTCAACCTGACCTGCGCGCGCGC----- 53963
QY 213 ArgPheGlyThrAlaThrValAlaIleGlySergThrValThrSergLysMetTrpPheSerg 232
Db 53964 -----GCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 54017
QY 233 GlyArgAspGlyAspLeuIleAsnAlaSerAsnAlaGlyThrValIleGlnAlaAspHis 252
Db 54018 GCGCAGCGCGCTATCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 54077
QY 253 AsnMetAspGlyAlaSer-----IleValIleGlnHisThrAsnGlyPheValSergSer 270
Db 54078 GGCCTGCGCGCTGACGCAACCTCATCATCTGATCAAGACACCAACCGCATCTTCTGCTGCGG 54137
QY 271 TyrIleHisIleLysAspAlaGlnValLysThrGlyAspThrValArgThrGlyGlnArg 290
Db 54138 TACGCGCACAACCAACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 54197
QY 291 IleAlaSerMetLysAsnGlnProSergLysAlaAlaLeuPheGlnIupheArgIleSergArg 310
Db 54198 ATCCCGAGATGGCGCCAGCGCAGCGCGCAGCGGTGCGCTGCGATCGAGTCCGCGGT 54257

```

QY 311 AsnGlyValIleValAspProLeuThrValIleu 321
Db 54258 CAGGCGACGCGGTTCATCCGGCGGAACCTGCTG 54290

RESULT 7
AB006073 3122 bp DNA linear BCT 17-APR-1999
LOCUS Pseudomonas tolaasii DNA for rpos sigma factor, complete cds.
DEFINITION AB006073
ACCESSION AB006073
VERSION AB006073.1 GI:2289783
KEYWORDS rpos sigma factor.
SOURCE Pseudomonas tolaasii (strain:PT814) DNA, clone:pHMH121.
ORGANISM Pseudomonas tolaasii
Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
Pseudomonas.

REFERENCE
AUTHORS 1 (sites)
Murai, H.
TITLE Characteristics of stress response in a mushroom-pathogenic bacterium, Pseudomonas tolaasii, during the interaction with Pleurotus ostreatus and carbon/nitrogen starvation in vitro
JOURNAL Mycoscience 40, 81-85 (1999)
REFERENCE 2 (bases 1 to 3122)
Murai, H.
TITLE Direct Submission
JOURNAL Submitted (26-JUL-1997) Hitoshi Murata, Forestry and Forest Products Research Institute, Bio-Resource Development; Matsumoto-1, Kuriizaki-machi, Ibaraki 305-8687, Japan (E-mail: muratoh@fpri.affrc.go.jp, Tel:0298-73-3211 (ex.458), Fax:0298-73-3795)

FEATURES
source
1..3122
/organism="Pseudomonas tolaasii"
/strain="PT814"
/db_xref="taxon:29442"
/clone="pHMH121"
1118..1123
/gene="rpos"
1132..2142
/gene="rpos"
1132..2142
/gene="rpos"
/codon_start=1
/transl_table=1
/product="rpos sigma factor"
/protein_id="BA21674.1"
/db_xref="GI:2289784"
/translation="MALSKAPEPDIDDEVLLMEILTEKSNDEGATPSVRTKS KNSTALRKHXYIDYRALDATQDLYNEIGSPILTPBEEVHFAALQKDPARKMT ESNLRLVVKIARVYVRGSLDLIEGNGLRAVEKPEPGRFSTYATWIMROT IERAIINQRTTIRLPIHVVELNVLRAARELQKLDHEPSPEINLLEKPYEVK MGINRVSVDVSLGPDSPKTLDTLDRPTDPCETLDDPLDQSSIDQWSELTKD QREVTIRRPGLRHSSTLEDVLEIGLTRERVRQIOVBGLKRLREIKNGLSSESL PQ"

BASE COUNT 698 a 845 c 897 g 682 t
ORIGIN

Alignment Scores:
Pred. No.: 3.18e-08 Length: 3122
Score: 286.00 Matches: 89
Percent Similarity: 44.13% Conservative: 50
Best Local Similarity: 28.25% Mismatches: 108
Query Match: 17.13% Indels: 68
DB: 1 Gaps: 13
US-10-018-706-2 (1-322) x AB006073 (1-3122)

QY 20 LeuIlePheGlyValIleThrThrCysIleLeuAlaGlyCysAlaSerIysProThrIyr 39
Db 235 CTGTGCTTGGCTTGTGATTCGCTATTTGGTGTGCTTCACGCTGCGCTGC--- 231

QY 40 AsnSerThrSerGlySerGlySerHisArgThrSerGlySerGlyIleuAlaIleGly 59
Db 292 -----AGTGGCGCCAGGCTGTGTCAGCCGACGTAACGCCGCGCTCCG 336

QY 60 SerGln---ValIleThrAspSerGlnGlyValProAsnArgIyrGlnValIleGlnGly 78
Db 337 CAAGCCGACGCTACACAGGCCA-----TACGTGCGCCGACAGGC 381

QY 79 AspThrValSerIleValIleAlaGlnArgIyrGlyLeuAsnTPArgIleGlyHisIle 98
Db 382 GACACATATGTTCTCGATCCGCTTCGTTACGTTGGGATTTACAAGCGCTCGACGCTGT 441

QY 99 AsnAsnLeuAsnSerSerIyrThrIleIyrThrGlyGlnIleThrLeuThrIleTPSerGly 118
Db 442 AACAAATATTCCTGTCGATATACATACATCCGGGCGAG----- 480

QY 119 AspLeuIleValArgIleuArgSerIleSerSerGlyValAsnThrAlaHisThrProSer 138
Db 481 -----ACCATCCGTTTGAGAGGGCGACGGGTTCCGGCCCGCA 519

QY 139 ProValAlaValGlnSerSerArgProProValGlnGlnHisProAlaValGlnIysPro 158
Db 520 ACCGTGTCGACACAGACCGGATCTTCGCG-----TCGCT 555

QY 159 ThrProProValValValValValIlysPro-----ThrProThrProProValValGln 176
Db 556 TCAGGCAAAACACCATCATCACCGCGCTGACAGCACCGCCGCTGGGTGCGAGC 615

QY 177 GlnProAlaProVal---AlaProProValThrGlnAlaProPheAlaThrGly----- 193
Db 616 AACCGGACACCGCGCGCTACCGCTGACAGGCCCGCTCCGACCGGTGGGGGTGCCA 675

QY 194 SerSerGlyValMetGlnPheArgIyrProValGlyAlaThrAsnProValValArg 213
Db 676 TCACACGCGGTGCTG-----ATTGAAA 699

QY 214 PheGlyThrAlaThrValAlaGlySerThrValThrSerAsnGlyMetThrPheSerGly 223
Db 700 TTC-----TCTTCAACCGTGTGTTGAATAAGGC 729

QY 234 ArgAsp-----GlyAspLeu-----IleAsnAlaSerAsnAlaGlyThrAlaIle 248
Db 730 ATTGATATCGCCGGAATTTGGACAGACCTGTTTACCTGCGCTGATGAGACAGTGT 789

QY 249 GlnAlaAspHisAsnMetAspGlyAlaSer-----IleValIleGlnHisThrAsnGly 266
Db 790 TACGCCGGAGGTGTTGGCGGCTACGGGAGCTGTATATCAACACAGCATACC 849

QY 267 PheValSerSerIyrIleHisIleIysAspAlaGlnValIysThrGlyAspThrValArg 286
Db 850 TACGTACAGCTTACCGCACCAACCTAGCTGTTGTTCCGGAGGCGACAGGTCAAA 909

QY 287 ThrGlyGlnAlaArgIleAlaSerMetIysAsnGlnProSerGlyValAlaLeuPheGluPhe 306
Db 910 GTCCGACAGCAATTCGCCAAAGGGGTCAACGGGTACACCGGGTGAACATGCACTT 969

QY 307 ArgIleSerArgAsnGlyValIyrValAspProLeuThrValIleu 321
Db 970 GAGATTGCGCGTCAAGGGAAGCTGTAGATCCGCTCAATTCTG 1014

RESULT 8
AF421351 5876 bp DNA linear BCT 07-MAY-2002
LOCUS Azotobacter vinelandii L-isoaspartate o-methyltransferase (pcn) gene, partial cds; muirkin endopeptidase (mep), stationary phase sigma factor (rpos), aerotaxis sensor (aerp), aerotaxis transducer (aert), and ferredoxin A (fda) genes, complete cds; and DNA mismatch repair protein (muts) gene, partial cds.
DEFINITION AF421351
ACCESSION AF421351.2 GI:20455801
VERSION AF421351
KEYWORDS Azotobacter vinelandii.
SOURCE Azotobacter vinelandii.
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae; Azotobacter.
REFERENCE 1 (bases 1 to 5876)

AUTHORS Kujat Choy, S.L., Meakins, D., Tindale, A.E. and Page, W.J.
TITLE The stationary phase sigma factor (rpos) of Azotobacter vinelandii
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 5876)
AUTHORS Whelan, H.L. and Page, W.J.
TITLE Aerotaxis in Azotobacter vinelandii
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 5876)
AUTHORS Tindale, A.E., Meakins, D., Kujat Choy, S.L. and Page, W.J.
TITLE Direct Submission
JOURNAL Submitted (19-SEP-2001) Biological Sciences, University of Alberta,
C4405 Biological Sciences Bldg., Edmonton, Alta T6G 2E9, Canada
REFERENCE 4 (bases 1 to 5876)
AUTHORS Whelan, H.L., Kujat Choy, S.L., Meakins, D. and Page, W.J.
TITLE Direct Submission
JOURNAL Submitted (06-MAY-2002) Biological Sciences, University of Alberta,
C4405 Biological Sciences Bldg., Edmonton, Alta T6G 2E9, Canada
REMARK Sequence update by submitter
COMMENT On May 6, 2002 this sequence version replaced gi:16226162.
FEATURES
source location/Qualifiers
1..5876
/organism="Azotobacter vinelandii"
/strain="WT"
/db_xref="taxon:354"
1..396
/gene="pcm"
1..396
/gene="pcm"
/codon_start=1
/transl_table=1
/product="L-isoaspartate O-methyltransferase"
/protein_id="AAL16089.1"
/db_xref="GI:16226163"
/translation="LEIGTSGYQTAVALQVERVESVERIQALQERAKERLELNIR
NVERFMGDGWEWGPALPYNGIWTAAASVPALDQLAPGRSLVIVSGAEQQL
LIVREPDGGRFRALETFRFVPLQGLF"
1033..1686
/gene="mep"
1033..1686
/gene="mep"
/note="similar to nlpD gene of Pseudomonas aeruginosa"
/codon_start=1
/transl_table=1
/product="murein endopeptidase"
/protein_id="AAL16090.1"
/db_xref="GI:16226164"
/translation="MRYGEMRELAHNGIAAPVVIHPGKIQIOLGPYRQAASTAAP
SVAPAPAPVPAATSOAAVAATPVKPSGKVPAPATKASAPVASRPSGMAPA
EGLVGPPSSNGSINKGIDIGGCGCPVLAASGVYAGGAGRGVCELIIHSDVY
VSAIGHNRRLVIRGQGVAKAGVIAEMGSTCTDRKLAHREIRGKRVDPDLQYLPFR"
1786..2790
/gene="rpos"
1786..2790
/gene="rpos"
/note="transcriptional regulator"
/codon_start=1
/transl_table=1
/product="stationary phase sigma factor"
/protein_id="AAL16091.1"
/db_xref="GI:16226165"
/translation="MALNIRAKKAPFVDVDDVLLMDADIVLEDDVLDDEPDLVPAKSR
QANLKQKVIDYTRALDOLYINIEISPSLLTPAEVHPARLQORDPGRKMMIS
NLRLVVKIARVYRGSLDLLEEGNLGLRAVEKDPERGFSTYATWIRQITL
RAIMQVTRIRLPIHVYELANVYLAARLELOKDHSPSAEILNLEKPAVEYKRM
GLNRSVTSVDSLPDSKTLVDVLTDRPDLDLQDDLSQISQIMQFLEDKRM
EVVVRFRGLRHSCTLEVGQETGLTRERVRQIYQVBAIKRLREILKNGISDALRQ
"

gene 3415..4356
/gene="aerp"
3415..4356
/gene="aerp"
/note="aerp"
/codon_start=1

gene 4435..5220
/gene="aerH"
4435..5220
/gene="aerH"
/note="aerH"
/codon_start=1
/transl_table=1
/product="aerotaxis transducer"
/protein_id="AAM22286.2"
/db_xref="GI:20501706"
/translation="MEQITATVQONNAANRAHSSISLHMAAEVQATGSAMKSVOSME
RISSSGKIAEIRIVDSIAFQTNILANASVEAARAGHGPFVAVSEVNNLATRS
AAADQVHALIEVSSREIHAGEOHVGAKEVAIROYEAVLEVDGIMSEIJAASERQTR
GIELVNOAVAKOMDEVTORNAELVOOSARDAANLDQVAVLSNVI SALLRVSGRGEHVS
RODRFAPARARRSEDPDRGKATVLEBGIAPPRERABF"
complement(5281..5604)
/gene="fdxA"
complement(5281..5604)
/gene="fdxA"
/note="fdxA"
/codon_start=1
/transl_table=1
/product="ferredoxin A"
/protein_id="AAM22287.1"
/db_xref="GI:20455804"
/translation="MAFVVDNICIKKTYDCEVCVPCDFYGRNFVVIHPREDICDA
LCERCPQALFSEDEVPEDWQEFITQNLNLAEMVNLITKQPLDPAEDMDGVGKGL
OHLER"
complement(5739..5876)
/gene="muts"
complement(5739..5876)
/gene="muts"
/note="muts"
/codon_start=1
/transl_table=1
/product="DNA mismatch repair protein"
/protein_id="AAM22288.1"
/db_xref="GI:20455805"
/translation="PQPFIQSDLPFASLPHPLMEELARKLPDDLSPQALELLYSWTR
L"
BASE COUNT 1040 a 1741 c 1923 g 1172 t
ORIGIN
Alignment Scores:
Pred. No.: 1,1e-07
Score: 282.00
Percent Similarity: 43.878
Best Local Similarity: 30.008
Query Match: 16.898
DB: 1
Gaps: 12
US-10-018-706-2 (1-322) x AF421351 (1-5876)
QY 23 GYVAILLETHRCYSILLESUVALAGLYCYALASERLYSPROTHRYTAAANSETRH 42
|||::|
::|::|
Db 892 GGGATCTCTCTTCCCTCGCTGGCGGCGTGGCCAGTTCCTCCGGAACAACGTGCAG 951
|||::|
::|::|
QY 43 SERGIYSERGIYSERHISARGTHRSERGIYSERGIYGLYENALAILLEGIYSERGINVAL 62

Db 951 -----
QY 63 ILETHRSAPSERGINGLYVALPRO-----ASNARGTYRGLNVALLYSGINGLY 78
::|::|
|||::|
|::|::|

Db 952 GTGGCGGAGCGCGAGTGGTCTCGGCGTACAGCGGCGAACAATGGTGCAGCGCTGGC 1011
 Qy AspThrValserIysIleAlaGlnArgTyrGlyLeuSerThrPargGluIleGlyHsIle 98
 Db 1012 GAATGCATCTCGTCATTCATTCAGCCGCTTATGGCTGGGAGTGGCGGCGAAGTGGCGCCAC 1071
 Qy AsnAsnLeuAnserSerIyrrHrIleTyrThrGlyGlnTrpLeuThrLeuTrpSerGly 118
 Db 1072 AATGGCATCGCTCGGCGCTCAATGATTCATCCGCGACAGAAATCCAGCTGGGG----- 1125
 Qy 119 AspleuLysValArgGlnSerIleSerSerGlyValSerThrAlaHsThrProSer 138
 Db 1126 -----CCGGTGGCGCGAGCGCGCGCTCG-----ACGGAGCGCGCGCGCTCA 1167
 Qy 139 ---ProValAlaValGlnSerSerArgProProValGlnGlnHsProAlaValGlnLys 157
 Db 1168 GTCCCGCGCGGAGCGCTCGCTCCAGCTGTCCAGCCAGCGAGCTGGCGTA----- 1221
 Qy 158 ProThrProProValValValValLysLysProThrProThrProProValValGlnGln 177
 Db 1222 -----GCGGCGACACCGGTCAAGCGGCGAGCTCGGAGAG 1254
 Qy 178 ProAlaProValAlaProProValThrGlnAlaProPhaAlaThrGlySerSerGlyVal 197
 Db 1255 GTTCAGCGCG-----CCGCTCGCGACAAAGCGCGCTGATGCA-----CAGCCGCGC 1299
 Qy 198 MetGlnPheArgTyrProValGly-----AlaThrAnProValAlaArgArg 213
 Db 1300 GTGTCCCGCGCGATCAGCTCGCGGTGGCGCTCGCGCGCGAGCGCTCGTGGGAGCGC 1359
 Qy 214 PheGlyThrAlaThrValAlaGlySerThrValThrSerAnGlyMetTrpPheSerGly 233
 Db 1360 TTTTTCCTCA-----AACGGAGT-----TTGAATAAGGCATGTGATCGCGGCT 1404
 Qy 234 ArgAspGlyAspLeuIleAsnAlaSerAsnAlaGlyThrValIleGlnAlaAspHsAsn 253
 Db 1405 CAATGGGCGCGAGCTGTCTCGCGCTCGATCGATCGGTGGTGTGTGTGTGTGTGTGTGTGT 1464
 Qy 254 MetAspGlyAlaSer-----IleValIleGlnHsThrAnGlyPheValSerSerTyr 271
 Db 1465 TTGGGGGGCTACCGCGCACTGATCATATCATAGATGATGATGATGATGATGATGATGATGAT 1524
 Qy 272 ILeHsIleLeuAspAlaGlnValValThrGlyAspThrValArgThrGlyGlnArgIle 291
 Db 1525 GGCCATACCGCAGCGCTGTGTGGGAGAGAGAACAGAGTGAAGGCGGCGGCGATATT 1584
 Qy 292 AlSerMetLysAnGlnProSerGlyAlaAlaLeuPheGlnPheArgIleSerArgAsn 311
 Db 1585 GCGCAATGGGCGTGCAGCGGTACCGATCGGAGTGAAGCTCATATTGCGCGCGCAG 1644
 Qy 312 GlyValTyrValAspProLeuThrValLeu 321
 Db 1645 GGAACCTGTGACCTGTGCAATATCTA 1674

RESULT 9
 AE003925/c 10026 bp DNA linear BCT 15-JUN-2001
 LOCUS Xylella fastidiosa 9asec, section 71 of 229 of the complete genome.
 DEFINITION AE003925 AE003849
 VERSION AE003925.1 GI:9105758
 KEYWORDS
 SOURCE Xylella fastidiosa 9asec.
 ORGANISM Xylella fastidiosa 9asec.
 Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
 Xylella.
 1 (bases 1 to 10026)
 Simpson,A.J., Reinach,F.C., Arruda,P., Abreu,F.A., Acencio,M.,
 Alvaranga,R., Alves,L.M.C., Araya,J.E., Baia,G.S., Baptista,C.S.,
 Barro,M.H., Bonaccorsi,E.D., Bordin,S., Bove,J.M., Briones,M.R.S.,
 Bueno,M.R., Camargo,A.A., Camargo,L.B., Carraro,D.M., Carrier,M.,
 Colauto,N.B., Colombo,C., Costa,F.F., Costa,W.C., Costa-Neto,C.M.,
 Coutinho,L.L., Cristofani,M., Dias-Neto,E., Docena,C., El-Dorry,H.,
 Facincani,A.P., Ferreira,A.J., Ferreira,V.C., Ferro,J.A.,
 Fraga,J.S., Franca,S.C., Franco,M.C., Frohme,M., Furlan,L.R.,
 Garmier,M., Goldman,G.H., Goldman,M.H., Gomes,S.L., Gruber,A.,
 Ho,P.L., Hobeisel,J.D., Junqueira,M.L., Kemper,E.L., Kitajima,J.P.,
 and Marino,C.L.
 The genome sequence of the plant pathogen Xylella fastidiosa. The
 Xylella fastidiosa Consortium of the Organization for Nucleotide
 Sequencing and Analysis
 Nature 406 (6792), 151-157 (2000)
 20365717
 10910347
 JOURNAL MEDLINE PUBLISHED
 REFERENCE
 AUTHORS
 2 (bases 1 to 10026)
 Simpson,A.J.G., Reinach,F.C., Arruda,P., Abreu,F.A., Acencio,M.,
 Alvaranga,R., Alves,L.M.C., Araya,J.E., Baia,G.S., Baptista,C.S.,
 Barro,M.H., Bonaccorsi,E.D., Bordin,S., Bove,J.M., Briones,M.R.S.,
 Bueno,M.R.P., Camargo,A.A., Camargo,L.B., Carraro,D.M.,
 Carrier,M., Colauto,N.B., Colombo,C., Costa,F.F., Costa,M.C.R.,
 Costa-Neto,C.M., Coutinho,L.L., Cristofani,M., Dias-Neto,E.,
 Docena,C., El-Dorry,H., Facincani,A.P., Ferreira,A.J.S.,
 Ferreira,V.C.A., Ferro,J.A., Fraga,J.S., Franca,S.C., Franco,M.C.,
 Frohme,M., Furlan,L.R., Garmier,M., Goldman,G.H., Goldman,M.H.S.,
 Gomes,S.L., Gruber,A., Ho,P.L., Hobeisel,J.D., Junqueira,M.L.,
 Kemper,E.L., Kitajima,J.P., Krieger,J.E., Kurame,E.E., Laigret,F.,
 Lambais,M.R., Leite,L.C.C., Lemos,E.G.M., Lemos,M.V.F., Lopes,S.A.,
 Lopes,C.R., Machado,J.A., Machado,M.A., Madeira,A.M.B.L.,
 Madeira,H.M.F., Marino,C.L., Marques,M.V., Martins,E.A.L.,
 Martins,E.M.F., Matsukuma,A.Y., Menck,C.F.M., Mitracca,E.C.,
 Miyaki,C.Y., Monteiro-Vitorello,C.B., Moon,D.H., Nagai,M.A.,
 Nascimento,A.L.T.O., Neto,L.E.S., Nhani,U.A., Nobrega,P.G.,
 Nunes,L.R., Oliveira,M.A., de Oliveira,M.C., de Oliveira,R.C.,
 Palmeiri,D.A., Paris,A., Peixoto,B.R., Pereira,G.A.G., Pereira
 Jr.,H.A., Pesquero,J.B., Quaggio,R.B., Roberto,P.G., Rodrigues,V.,
 de M. Rosa,A.J., de Rosa Jr.,V.E., de Sa,R.G., Santelli,R.V.,
 Sawasaki,H.B., da Silva,A.C.R., da Silva,F.R., da Silva,A.M., Silva
 Jr.,M.A., da Silveira,J.F., Silveira,M.L.Z., Siqueira,W.U., de
 Souza,A.A., de Souza,A.P., Terenzi,M.F., Trufelli,D., Tsai,S.M.,
 Tsuchino,M.H., Vallada,H., Van Sluys,M.A., Verjovski-Almeida,S.,
 Vettore,A.L., Zago,M.A., Zatz,M., Mendes,J. and Setubal,J.C.
 Direct Submission
 Submitted (02-JUN-2000) Organization for Nucleotide Sequencing and
 Analysis, Bioinformatics Lab - IC/Unicamp, C.P. 6176, Campinas, SP
 13083-970, Brazil
 Location/Qualifiers
 1. 10026
 /organism="Xylella fastidiosa 9asec"
 /db_xref="taxon:160492"
 /clone="9asec"
 86. 1393
 /gene="Xf0851"
 86. 1393
 /note="similar to SP|P29011 (percent identity: 54 %/query
 alignment coverage: 95.6 %/subject alignment coverage:
 96.3 %) ; identified by sequence similarity; putative; ORF
 located using Glimmer/RBSfinder/Start codon shift: -6"
 /codon_start=1
 /transl_table=11
 /product="D-amino acid dehydrogenase subunit"
 /protein_id="AAF83661.1"
 /db_xref="GI:9105759"
 /translation="MRVILIGSGVITGTSAMWLSKAGCEVVVDRSGAGLETSYANG
 GOLFSTPMAPGVPLKAVRWLFEHRAPLSRPTDMNQWYLMARHCSAERA
 VNKSRLRLSEYREALBALSAFTGTPGRLHGTQLPRTQOGLVAVADILROY
 GIPYEVLSPOSKSPKRGDAGSVRPGRLRLPHDQGCCCTPTQGLAARGVF
 RYCTYQRLVDEGPRVTGMINALERADCCVVALGSYSFELIAPLGLRLPYPLGFG
 SLTLPMIDASRAVSTVLDESXYAVATRDERTIRVAGMAVSGDVSLNERRATLEM
 VWDVYVGGDLDGGEFWTGLRPATPDGIVGATGYOGFLMTGTGLTGMWSSSG
 RYADIDICRPPCISSESGDMFYIVSTIPQECAPCVPTP"
 1363. 2460
 /gene="Xf0852"
 1363. 2460
 /gene="Xf0852"
 /note="similar to SP|P06655 (percent identity: 45 %/query
 alignment coverage: 98.4 %/subject alignment coverage:

[illegible]

gene
CDS
/gene="tal.1"
/codon_start=1
/transl_table=11
/product="tal"
/protein_id="AAK03686.1"
/db_xref="GI:12721999"
/translation="MTTQDLSRMNTVVAVDTGDIIEAIKKYQPEDATTNPSSLISASAL
LPOVALSDEKAIATAKSKNSCKOOLIDAEPKLAIVNIGLEIKTVLPGRISREVPARLS
YDTQATTEKAKKLIALAKYNSISNDRLIKIASWQIGRBAEELKQGINNTLLPRLS
EAQAAACEAGAVYLISPPVGRILDMYKRNPSKYAPAEEDGVISVKTIINYQDHGR
NTIYWGASFNNVGETIELAGCDRLITPAALKEIQENNAPEERKLSYGEKAPQPLP
TEAFERYQNSNDNAWEKLDADGIRKTFPAADQEKLEAMLLTKF"
complement(1254. .2081)
/gene="PM1603"
complement(1254. .2081)
/gene="PM1603"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAK03687.1"
/db_xref="GI:12722000"
/translation="MKYAFIDYENIHSLDININLKRYKLYFLGAAQOOSITLSKEND
QIVHLLITKIVAKNNLDLPHLAYFLGKFDQCVDNKIDVLSKQKQIGCIDLQONOR
NARNQOLLTKLKSTADKOPSEVKEKEKELDAQONKAYELIKTLKNIISTDLSKLE
TLSNHSBQDTHKHDPDLSVITIDKYLAKETPEIIOINNTIYKRVSLPIYLS
YLLKKQKKRPTKLSLKNDAISKLNKKNNEVDKIIDLPRANNIILNTQAEKQYK"
complement(2102. .2554)
/gene="PM1604"
complement(2102. .2554)
/gene="PM1604"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAK03688.1"
/db_xref="GI:12722001"
/translation="MHKKNVTALCVYHCKGKFLFVELEHGKLTILNOPAGHLADESIL
LEGARKEIFERTGTHADMONLIKTYQWHAAPSCKDPLRFTPAVSLDMLPIQPODPDIT
TRGLMLSLSEFYDYIQAGQOCERSPLVESYDLYLKGEFISVQLKAF"
complement(2567. .3292)
/gene="PM1605"
complement(2567. .3292)
/gene="PM1605"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAK03689.1"
/db_xref="GI:12722002"
/translation="MHKPTKNOCKTPRPSLSRSNKRKAKEIKQRPLSFRKQTV
NPAEYVLLENKPYDVLTFQSDSEGRATLKAFITIPHYRGRILDRQSEGLVLTNNG
EIQHLLAERKQNTKTEYTPAAYQEGIPQETDLAQVLQGVLANDGKLPKAVRLVPPNEP
WQNRNPIERKQITPTSWLEIKIHGSRNQVRRMTAHIGFTLRILIRYQMLFNLANLE
QGOYVQSEVEELTALYQQLKLFK"
3481. 4728
/gene="tdp"
/note="PM1606"
3481. 4728
/gene="tdp"
/codon_start=1
/transl_table=11
/product="tdp"
/protein_id="AAK03690.1"
/db_xref="GI:12722003"
/translation="MQSKVVIPOEKKIQLDAKQALIVDPHPILPIEGDGIQVDTVPAR
MRVTDIAVEKAYGGRKISWMEIYAGKANVEYGEWNTLPDEMTFTRLDVNAIKGEB
LMTPEGGISRLVAMROGDLVNCRLPIRYDGPSPVYGRBELVDVITPRENSDAIY
AGVEVVASAANKYIAFLQOEMGVKKIRFPEDDGIGIKPKSKGTORLPRALQYVYI
DNDRKSLTLVHKGNIMKFTBGAFKEMGVOVAQBFAGELDDGPMWTLKNPTGSLIIL
KDSLADAPLQSLILHPTEYDIATLNINGDYISDALAAOVGGIGISGANINGEAALEIF
EATGHTAKRIGDQKNGPGLISLNGEMLRHLGWLLEAADIIVKAVSKTIADKTVTFPPF
AEMLEGAALRSTSEPADQIVANM"
4885. 5163
/gene="PM1607"
4885. 5163

gene
CDS
/gene="PM1607"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAK03691.1"
/db_xref="GI:12722004"
/translation="MRLFIPLIVATLLFOYDFWNGKGYLDYKRTAQALQHKQENE
KLSCHNOVAALAIKILKQGVLEIRERARFQIDMWKPEELFIHIVKQK"
5163 .5879
/gene="PM1608"
/gene="PM1608"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAK03692.1"
/db_xref="GI:12722005"
/translation="MTPIFAKSTRIRIVAVVPAAGIGSRMOKDKKQYLIHKITLH
TSLVILGLPELRIKILIAVAANDPYITGCLLTHPKIQLVSGSSRASPVLGNLAVKS
AVQNSQEDFWVHDARPCQLTHQIDLDKLVVEDQNGALIAIPDITIKRALINQIHY
TQSDNFWAQWQPPQFPFIATLQAALQALQGLQVTDASAMEFAGRPRLVAGRSNDN
IVTRPEEFLALFYLSTRK"
5888 .6364
/gene="PM1609"
/gene="PM1609"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAK03693.1"
/db_xref="GI:12722006"
/translation="MIRIGHGPDVHAFGGEPIIIGVAIPYKGLLAHSDGDVALHA
LDLALGALVALDIGLTFPDITMOYKGRASRLLEAVTQVAKYKGVKATVLIIOQ
AKMRPHIDMAAALAEADLACDIEQVNVKATTSERLTGFTGREGIACAAVALVKS"
6364 .7371
/gene="PM1610"
/gene="PM1610"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAK03694.1"
/db_xref="GI:12722007"
/translation="MELALVTRRPEQYARLAKQCDPIYKENTLYMSGDSEFPAIVKY
RKDTCUATVHGKLAQFVGISERNNGYGLKQKRAVTFQWPCILHMGQPTDFRFSFO
EGVDILIEVTRHRIKIRTGSLRGNHFEILRGAKEDELVNRIKNTIKQGGFNITFEQD
FGDGNILVQALRMAQGEINVDKRRKRSFYSAASEVFNILVNSERIALQLAQQVLRK
DMQLQGSMSFQADEKEDLNALQRLRQDDTLTPAPIGEONPATDIEQLVQKQIC
ALITLWAKRMRMAARRPMLMQAQLQMAFVABGLKLAFLPAGSVAITALVEVNVQDE
D"
gene
CDS
/gene="PM1611"
/gene="PM1611"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAK03695.1"
/db_xref="GI:12722008"
/translation="MORMSKTLTSLILVSTMNVAFOREEDPVFQGVFAVYKRGDHA
AFPLTMSRAQDEMSQFNVGMYDDGVEDVDKQOALKWYKSAEONHPDAQYHGLG
MYSEGGPIADQEPKQAYKMYSQSAVGGDADGALNLIGITLVANGSIEIRMDRBAKMYFKQ
CKAGLPEGGDME"
7991 .8731
/gene="PM1612"
/gene="PM1612"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAK03696.1"
/db_xref="GI:12722009"
/db_xref="GI:12722009"

gene
8771..9346
/gene="lppB"
/note="Pm1613"
8771..9346
/gene="lppB"
/codon_start=1
/transl_table=1
/product="lppB"
/protein_id="AAK03697.1"
/db_xref="GI:12722010"
/translation="MKIIGAMYDKTMAKSKRRVATFLATLSFTLEAFPIPPDVMTL
PMSISQPORALKAFYVAASVIGGVGIGYAAVAVFVHYVQDFQYQHWETAFISM
FERKGIIVAVAGSGSPIPYKVFPTICACVMOMFLPFLVTAISPAKFLIVAKLAAGN
GEEKPADLKRSIEILIGMSVVLIAVIAYLILR"
9559..10962
/gene="Pm1614"
9559..10962

Alignment Scores:

Pred. No.:	2,21e-05	Length:	12956
Score:	49.50	Matches:	80
Percent Similarity:	45.69%	Conservative:	42
Best Local Similarity:	29.96%	Mismatches:	110
Query Match:	14.94%	Indels:	35
DB:	1	Gaps:	9

US-10-018-706-2 (1-322) X AE006198 (1-12956)

73 TyrGlnValIysGlnGlyAspThrValSerLysIleAlaGlnArgTyrGlyLeuAsnTrp 92

Db 10201 TATACGTCGTAAGCGATACAATGTACTCATCGCCTATATTCAAGCTTAGATGTC 10260

93 ArgGluIleGlyHisIleAsnAsnLeuAsnSerSerTyrThrIleTyrThrGlyGlnTrp 112

Db ..|||..
10261 AAAGATTGGCATCATTAAACAATATGTCCTGAACCTTATCGTCTAAGTGTTGGCAACC
10320

113 LeuThrLeuTrpSerGlyAspLeuLysValArgGluArgSerIleSerSerGlyValAsn 132

Db 10321 TTACGTGTGTCGATGGA-----CGTGTGGCAAGCACCTCTTCACAACCACTG 10361

133 ThrAlaHisThrProSerProValAlaValGlnSerSerArgProProValGlnGlnHis 152

Db 10369 ACA-----CAACCTGTGACGGTACCCGTTTCAACAACCG-----AAGAGCAGT 10414

153 ProAlaValGlnLysProThrProPro-----Val 162

Db 10411 GAAGTAACCTATACCCCGGTCACATGGTACACACATACGGTTCGGATGTACCATCAT 10470

OV 163 ValValValLysLysProThrProThrProProValValGlnGlnProAlaProValAla 182

Db 10471 GGACCAATTAATCAGGTGTCAGTTCGCCCTGTTCCAGTCCCAACCTGAACCAGTGTT 10530

183 ProProVal-----ThrGluAlaProPheAlaThrGlySerSerGlyValMet----- 198

Db 10531 AAGCTGTTGAAGCAGCTCTGTACCTGTGCCAAGCAGCAGTAACACATGTTCTG 10591

199 -----GlnpheArqTvrProValGlyAlaThrAsnProValValArqArpPheGlyThr 216

Db 10591 AATGTACTTGGCAATGGCCAACCAAGCAAT-----ATTGTACAAGGCTTTCACC 1064

217 AlaThrValAlaGlySerThrValThrSerAsnGlyMetTyrPheSerGlyValArgAspGly 236

10645 CAGACCGTGC-----AATAAGGATTGATTGCTGGCTACCTGGT 10668
 ||| :: ||| :: |||
 || :: ||| :: |||

237 AshtleuTlaSnA]aG]vThrvA]T]eG]nA]aAspHisAsnMetAspGly 256

1 0680 CAGCCGACAGGAGTATGCGGTAATGCATTACGTGCT 1074

```

QY 257 AlaSer-----1LeVal11IleGlnHisIThrSenglyPheValSerSerTyT11HisIle 274
Db 10750 TATGGCACTTAATATCATTAATAACACATCATATGACTTATTAAGTGCCTATGATCCAAAT 10800
QY 275 LysAspAlaGlnValIlyeThrGlyAspThrValArgThrIlyGlnArg11LeAlaSerMet 294
Db 10810 GAAAGTATTTTAACTGAAGAATCAGACAGAAAGTGTGCGGGTCAACAGATTGCCAAATG 10866
QY 295 LysAsnGlnProSerGlyAlaAlaLeuPheGluPheArgIleSerArgAsnGlyValTyr 314
Db 10870 GSTAGCTGTGATGCCAATATAGTGTCAAACTTGACTTTGAATTCCGTATTAAGCAATCA 10922
QY 315 ValAspProLeuThrValIleu 321
Db 10930 GTTGATCCACCCGTTATTTG 10950

RESULT 11
PPY19122
LOCUS PPY19122 3539 bp DNA linear BCT 07-JAN-2000
DEFINITION Pseudomonas putida nlpd (partial), rpos, fdx, muts (partial)
genes.
ACCESSION Y19122
VERSION Y19122
KEYWORDS fdx gene; ferredoxin; lipoprotein; mismatch binding protein; muts
SOURCE gene; nlpd gene; RNA polymerase sigma factor; rpos gene.
ORGANISM Pseudomonas putida
Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
Pseudomonas.
1 (bases 1 to 3539)
Kojic,M., Degraessl,G. and Venturi,V.
Cloning and characterisation of the rpos gene from plant
growth-promoting Pseudomonas putida WCS358: Rpos is not involved in
siderophore and homoserine lactone production
Bichim. Biophys. Acta 1489 (2-3), 413-420 (1999)
20135559
MEDLINE 10673044
PUBMED 2 (bases 1 to 3539)
10673044
REFERENCE Venturi,V.
AUTHORS Direct Submission
TITLE Submitted (24-JUN-1999) V. Venturi, International Centre for
JOURNAL Genetic Engineering & Biotechnology, Bacteriology Group, Padriciano
99, 34012 Trieste, ITALY
FEATURES
source location/Qualifiers
1..3539
/organism="Pseudomonas putida"
/strain="WCS358"
/db_xref="taxon:303"
1..735
/gene="rpos"
<1..735
/gene="nlpd"
/codon_start=1
/transl_table=11
/product="lipoprotein"
/protein_id="CAB46190.1"
/db_xref="GI:5305129"
/db_xref="SPTRMBL:Q9WV6"
/translation="KLVLALAWGSLLAGSGSTSGTSARVDNRNNAKRPPTYSGGY
IVKGDVLFSLAFRGMVDEKELARNNGIPAYTIRPQPIRFSSGSGSTTVSSPSS
SSKTVIRPVRGASGASPASTTKPATPTIPAVVAIVPAEBAVGMTWPAANGVLI
GKPLANSGLINKIDIDAGLQGPVPAASGAAVVYDGGIRGVAELIIKHSDTVYSAVG
HNRLRLVREGQVVAAGOSIAEWVGH"
920..1927
/gene="rpos"
920..1927
/gene="rpos"
/gene="rpos"
/codon_start=1
/transl_table=11
/product="rna polymerase sigma factor"
/protein_id="CAB46191.1"
/db_xref="GI:5305130"
/db_xref="SPTRMBL:Q9WV5"

```

	gene	complement (2429. .2752)					
	/gene="fdx"						
CDS	complement (2429. .2752)						
	/gene="fdx"						
	/codon_start=1						
	/transl_table=1						
	/product="ferrioxin"						
	/protein_id="CA846192.1"						
	/db_xref="GI:5305131"						
	/translation="MPFVYTDNCIKCKYVDCEVCPCDPEEGPNFLVIHPDCICDA LCBECPPQAQAIPISEDEVPSGMENFIEMAEIAEWITERRKALPDAEWMDGKTGT ADLER"						
gene	complement (2873. .3539)						
	/gene="muts"						
CDS	complement (2873. .3539)						
	/gene="muts"						
	/codon_start=2						
	/transl_table=1						
	/product="mismatch binding protein"						
	/protein_id="CA846193.1"						
	/db_xref="GI:5305133"						
	/db_xref="SPTREMBL:O9MWY4"						
	/translation="SFVPAQCCELSLDVRIIFRIGSSDDLAGRSFTFMESNTANIL HNADRSLVMEDEVRSTPFDGISLMAAEBRLAQRAVTLFAHYAFLEPVASESP LVANVRLNATEHMERIVFLHAVLPGPSOSGYGAVALQAVPAVIVORAHEHGRLT ASLPHQEPHSHTAKADLAHVHQSDLFASLPHPAIEKLGKDLDMDFPQAIEMLYQAK NIL"						
BASE COUNT	678 a 1006 c 1098 g 757 t						
ORIGIN							
Alignment Scores:							
Pred. No.:	9.74e-06	Length:	3539				
Score:	245.00	Matches:	95				
Percent Similarity:	43.59%	Conservative:	41				
Best Local Similarity:	30.45%	Mismatches:	119				
Query Match:	14.67%	Indels:	57				
DB:	1	Gaps:	14				
US-10-018-706-2 (1-322) x PPR19122 (1-3559)							
Qy	20 LeuilephegiValiiethrThCyisileuAlaglyCysalaSerIysProthrTyr	39					
	::: :::						
Db	7 CTGGTGCTTGACATCGGCATCGGCACCTTGCGCGGTTGC-----	48					
Qy	40 AsmerThrserglySerglySerHisArgThrserglySerglyIleuAlallegly	59					
	49 ---TCACAGCACACTGCACAGAGCCGGGTGTGCCCTTAACAATGACGCCCAAG	105					
Db	60 SerGlnValIIethrRapserGInglyValProAsnargTyrgInValIylsGInglyASP	79					
Qy	106 CGCGCCGACGGTACCTCCGGGCAA-----TACATCGTCAAGCCGGGCGAT	150					
Db	80 ThrValSerIysIleIagInargTyrgIlyLeuAsnTrpargIuIlegIyHisIleasn	99					
Qy	151 ACGCTGTTCTCATCGCTTCGGTTACGGCTGGGACTAACAAAGAGTTGGCTGCCCGTAAC	210					
Db	100 AsnleuAsnseryrrThrlleTyrrThglYcInTrpleuthrLeutrpSerglyASP	119					
Qy	211 GGCATCCCGGGCGCTTATACCATTCCGCCCGGCCAGCGCATTCGATTCAACAGCGGT---	267					
Db	120 LeuIysValArgIuaYgserIleSerSerGlyValasnthr---AlaHisThrProSer	138					
Qy	268 -----TTCACGGGACGACCAACCGGTGGTGTCCAAGCCCTTCG	303					
Db	139 ProValAlaValGInserSerArgProProValGInGlnHisProAlaValGInIlyPro	158					


```

/protein_id="AAC75890.1"
/db_xref="GI:1789215"
/translation="MSTETIEFNNSEDMANOLGALSGENLALLHGLTPDILRIY
AYADYHEKGNITDAEIVYKELCIYAFENHEHYLLNDPASPCKKKYQOAAVLYLSTN
YPPDYDSVYIRMGCCQIGAKNIDNNAQCFFHITNNCEDSVSKAQAYIELNDNSE
DNG"
protein_bind
1391..1408
/notes="central position to predicted promoter: -208"
/bound_molecety="Phob predicted site"
promoter
1575..1602
/notes="factor Sigma70; predicted +1 start at 2990069"
1656..3032
/genes="ygeH"
/notes="b2852"
1656..3032
/genes="ygeH"
/notes="ygeH"
/function="putative factor; Not classified"
/notes="o458; This 458 aa ORF is 28 pct identical (6 gaps)
to 407 residues of an approx. 560 aa protein IAGA_SALTI
SW: P43016"
/codon_start=1
/transl_table=1
/product="putative invasion protein"
/protein_id="AAC75891.1"
/db_xref="GI:1789216"
/translation="MDLENKFSYHLEGLTLTEDGILTOGNEQYVPOKEGLVILV
ESAGVLLKMLISRWKNIIIVSDSEILTRCIYSKCIPEKIGYGRCIETTYKGYRS
GQVKTINEDNTSDYSIAIFPTTSANTDPLINQELVQIISNKIDGITYPMMA
TNPNDHISQNSPLSRFPDYFVGRINQNAVATLIELIDAKNLPLIASNHLPIVE
LHNTSOFIINDILOTVAHKPEKSVRLAKQDQYKNHYLSDEMLAKKELYPETPSIYR
AMTIFRLQNSKSDIOTLKTECYCLACHNSIALHGSELELAQKALELDDYSDIT
TVDKIILAMGLITGLSGQAKVSHILPEQAKIHSITDASLYYVALVHPNEKEEAR
ICIDKSIQLEPRRRKAVIVKECVDMYVNPINKNIKLYKETSESHRVIINDLIKIK
QUTRICMR"
3039..3061
/notes="central position to predicted promoter: -100.5"
/bound_molecety="Fur predicted site"
3039..3061
/notes="central position to predicted promoter: -113.5"
/bound_molecety="Fur predicted site"
3118..3145
/notes="factor Sigma70; predicted +1 start at 2991612"
3130..3158
/notes="factor Sigma70; predicted +1 start at 2991625"
3200..3418
/genes="b2853"
3200..3418
/genes="b2853"
/function="orf; Unknown"
/notes="o72; This 72 aa ORF is 33 pct identical (1 gap)
to 54 residues of an approx. 2520 aa protein TUD_DROME SW:
P25823"
/codon_start=1
/transl_table=1
/product="orf, hypothetical protein"
/protein_id="AAC75892.1"
/db_xref="GI:1789217"
/translation="WTFPIGINLSGSNNANATGDEVSLDKHINTSATDIOQAF
IVSTWAPFQNDMYSBPNPISPYKIEW"
3521..3549
/notes="factor Sigma70; predicted +1 start at 2992016"
3561..3977
/genes="b2854"
3561..3977
/genes="b2854"
/function="orf; Unknown"
/notes="o138; This 138 aa ORF is 39 pct identical (7 gaps)
to 128 residues of an approx. 168 aa protein IAGB_SALTI
SW: P43018"
/codon_start=1
/transl_table=1
/product="orf, hypothetical protein"
/protein_id="AAC75893.1"

```

```

/db_xref="GI:1789218"
/translation="MSFASNSCSNAGTWERIEPNLIKATALVESNLKQDSIGNRDK
NNNIKSLDYLMQNHILPLKRGIIKERDILNPCLNIKIETIILVNHRSRCGV
TWOCITGVYAGFPMQNDQKROOYAPKILVILPOLMN"
complement (4022..4468)
/genes="ygeK"
/notes="b2855"
complement (4022..4468)
/genes="ygeK"
/function="putative regulator; Not classified"
/notes="f148"
/codon_start=1
/transl_table=1
/product="putative 2-component transcriptional regulator"
/protein_id="AAC75894.1"
/db_xref="GI:1789219"
/translation="TWGAELVKVVKSHKIDAHITTVAKKPYDSIKLEPAKGCW
KTHSPKALNDAIDISNGYTPDPSVHNDCKRISRSNDQITNREELIQLADGKT
NKEIAPFLQLSRKTVEYTHRLINKLDVHSGIELIKTALMGVCTI"
complement (4499..4654)
/genes="b2856"
gene
Alignment Scores:
Pred. No.: 7,71e-05 Length: 9720
Score: 238.00 Matches: 78
Percent Similarity: 39.62% Conservative: 46
Best local Similarity: 24.92% Mismatches: 103
Query Match: 14.25% Indels: 86
DB: 1 Gaps: 8
US-10-018-706-2 (1-322) x AE000369 (1-9720)
QY 16 LysArgLeuGlyLeuIlePheGlyValIleThrThrcysIleLeuAlaGlyCysAlaSer 35
Db 9429 AATCTCTGGATTCGGATGATGTTATTCGGTTGAGCTCTTTGGGGGCTGT----- 9376
QY 36 LysProThrTyranSerThrSerGlySerGlySerHisArgThrSerGlySerGly 55
Db 9375 -----TCGGGTGCAAAATCATCCGATCAGACGAGAGCTATTCGGGC 9337
QY 56 LeuAlaIleGlySerGlnValIleThrAspSerGlnGlyValProAsnArgTyrglnVal 75
Db 9336 -----TCGGTTTACACCGTG 9322
QY 76 LysGlnGlyAspThrValSerIysIleAlaGlnArgTyrglyLeuAsnTyrpArgIle 95
Db 9321 AAACGGGGGATACGCTATATTCGATTTCGGGACACGAGGAAACGCTG 9262
QY 96 GlyHisIleAsnAsnLeuAsnSerSerTyThrIleTyThrGlyGlnTyrPleuThrLeu 115
Db 9261 GCGGAGCTGAACGCGCATTTCCCTCCCTTACACCATTAAGTTGTCGAAACTTAAACTG 9202
QY 116 TrpSerGlyAspLeuIys-----ValArgIleuArgSerIleSerSerGlyVal 131
Db 9201 -----GTTGGGGGGAAGATAGACGATTTACAGTAAATCAACCGCAATCAAGACC 9148
QY 132 AsnThrIleHis---ThrProSerProValAlaValGlnSerSerArgProProValGln 150
Db 9147 AAACCGCATCGCTTACACCGCTACGCGGTCACGAAATCATCTCGCCGCGCTGGGG 9088
QY 151 GlnHisProAlaValGlnIysProThrProProValAlaValIysIysProThrPro 170
Db 9087 CAACGTTGTGTTATGATGCGACACGACGAGAAATTTATCATG----- 9046
QY 171 ThrProProValValGlnGlnProAlaProValAlaProValThrGluAlaProPhe 190
Db 9045 -----CCGTAT 9040
QY 191 AlaThrGlySerSerGlyValMetGlnPheArgTyrrProValGlyAlaThrAsnProVal 210
Db 9039 TCGACACGACATGCGCGC----- 9022
QY 211 ValArgArgPheGlyThrAlaThrValAlaGlySerThrValThrSerAsnGlyMetTrp 230

```

Db	9021	-----	::::	AAATTAAGGATTTCAT	::::	9007
Qy	231	Phesercglvargapglvyspleuileasna	laser	naalaglythvalilegla	250	
Db	9006	ATTCACAGTCACCGGGGTACCCATATTAC	CCCCCGGGGTGACGAGAAAGTGTGATCTG	8947		
Qy	251	Asphisasnmelaspglvylaser	-----	llevalileglnlstharsnglypheval	268	
Db	8946	GGCACACCGAGCGCGTGGCTACGTATCTCATGATTAACACAGAAAGATTACAT	8887			
Qy	269	SerserTYTllehisilelyaspaileglnvallysthrglyaspthrvalargthrgly	288			
Db	8886	ACGGCTTACGCCCATATATGACACGATGCTGTGTAATTAATGGCCAAAGGATGAGGCTGG	882			
Qy	289	qlnarTllealasermelaysnclnproserglyalalaleupheclupheargile	308			
Db	8826	CAAAAAATCGGCACATATGGGAGGACCGAGATGGCGCATCTGTTCCCTGCATTTCCAGATT	8766			
Qy	309	SerargasnglyvalTYvalaspProleuthrvalleu	321			
Db	8766	CGTAAACGTGCACACCGCAATTGATCCGCTACCTTACTCTG	8728			

ECU28375/c	55175 bp	DNA	linear	BCT 08-DEC-1995
ECU28375				
Escheichia coli K-12 genome; approximately 64 to 65 minutes.				
LOCUS				
DEFINITION				
RESULT 13				

ORGANISM	<i>Escherichia coli</i>
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;	
<i>Escherichia</i> .	
REFERENCE	1 (bases 1 to 55175)
AUTHORS	Plunkett,G.
TITLE	Direct Submission
JOURNAL	Submitted (01-JUN-1995) Guy Plunkett, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
COMMENT	This sequence was determined as part of the E. coli Genome Proj ect directed at the University of Wisconsin.

Testation/Organizational

```

source
1. .55175
/organism="Escherichia coli"
/strain="K-12"
/db_xref="taxon:562"
/mep="approximately 64 to 65 minutes"
/note="this sequence comprises the following lambda
clones: DD644 (EC30K4.64-1), DD645 (EC27-1154),
DD649 (BC22-159), and DD653 (EC18-341); M13Janus vector was
used for subcloning"
370. .816

```

```

/db xref="GI:887802"
/translation="MSRETETIEFNNSDEMANOLKHALSKENALALHGLTPDILDIRIY
AYADYHEKGNITDAETIYKFLCIYAEHNHEYLKDFPVSOCPPKKYQOADAIVKLSYN
YFPDDDSVYIRWMOQCIQIANKINDNACFYHIINCCEDDSVSKAOAYIELINDNSE
DNG"
1910, 3266
/protein="ORF_0458"

```

/translaction="m'dlenkrsyafelsclt'etbegit'loqneocvayit'pokenkayivall
 esbaachyaxkdmitswknin'lydespelt'rcys'p'rcit'pektgyrcv'it'rkayrks
 covp'rtk'ntent'sd'sy'la'p'p'p'tt's'nt'nt'dp'lin'el'no'v'la'it'ink'rid'g'ly'p'ma'na
 t'n'rcnd'h's'ons'f's'r'ep'p'p'p'v'ng'r'inn'onna'v'nt'ly'el'da'gn'l'f'i'as'n'h'l'p'v'de
 l'ant's'of'i't'd'it'it'ly'p'p'p'p'v'la'rod'q'p'v'la'k'h's'el'ak'kel'y'd'ep't'p's'i'y'r'a
 l'ant'i'd'r'q'ns'd'it'ot'it'k'ec'y'v'la'c'h'm'e'l'ah'g's'e'l'a'k'e'l'ad'y'd'it'it'
 t'v'd'g'k'it'la'p'w'it'it'g's'o'a'v's'h'i'l'p'eo'a'c'h't's'd'it'la's'y'y'a'l'v'p'h'ne'k'i'ea'a'r
 i'c'ib's'd'p'el'err'r'a'v'i'k'ec'y'd'm'y'p'n'p'l'k'nn'i'k'ly'k'et'ee's'h'r'v'id'it'it'l'k'k'j'
 q'it'r'i'c'k'v'c'

```

/translation="MTNPIGINNLSQSSNIANAATGDEVVSLDKHINTSATPTDQIQAH
IVSTWMPQNDMYSEDNPISPYKIEI"
3816.4232

```

```

/protein_id="AAA83036.1"
/db_xref="GI:867805"
/translation="MSFASNSCSNAGTMTFRLEPNL IKAIALVESNLKKDSIGKQRDI
NNNISISLDVLMQINQMHIPLTKRGGIKIDEDDLNDNCLNKIGTEILLYNHFRCGG
TMOCGTINAGFAMNKKKBOOYAPKYLILYIGLNN"
complement(4277..4723)

```

```

/protein_id="AAA83037.1"
/db_xref="GI:887806"
/cranialation="MMGABLVKRVKSHKIDAHITFFVAKMPYIDSIKILBAGAKCCV
KSHPKAKIRADISINQGYTPDSVHMDCKISRSRNSQNLNTESELIQIADQK
NKEIAFLQSLKRYTETHRLNIMKLDVHSGELIKTALRMVCVCI"
complement(4754..4909)

```

```

/label=EUKT_L21
/protein_id="AAA83038.1"
/db_xref="GI:887807"
/translation="MGKIKIVSDQPFMIDGILGHPDLYEWGSKDLKKAIECKNSTA"
complement (5131..5562)
/note="ORF #143"

```

```

/protein_id="AAAB3039.1."
/db_xref="GI:887808"
/translation="MINDLSILIKSEEDVYIKIFESWTKLPSISGVNHIPT
FKKSLFVEYIPVSKINWVHIFEDHDKFVFPTDQFIASFQEVNDCEQLINN
CFSPDKVNICEQARRHYLEKMPETHSLDNNNSVLA$PDL"
complement(5565..5837)
/notes="WOP 590"

```

```

CDS
    /label=ORF_f90
    /protein_id="AAAB3040.1"
    /db_xref="GI:887809"
    /translation="MRKTIEMSLKSPANGVIKRSISDGLKEIVLSKEKILLETAK
    IOSIERKREKRIQGYDGYTGIIIDEMDNFIPILSLCSELEKKE"
    complement(5779..6204)
    /note="ORF_f141"
    /codon_start=1
    /label_table=11
    /protein_id="AAAB3041.1"
    /db_xref="GI:887810"
    /translation="MSGNIGANPINNNMLPLCLISGCHFYRERFAERGFFYKVDV
    LBDYSAIPLAEINERKARYKGIANYNIITCGSTLLPYRPOQPLAMQORFNLFPDF
    VDHIOPLPLASFLERITFYAKXNDDELKISCKKCD"
    6178..7508
    /note="corresponds to V00610 (ISIS2) 1..1327"
    complement(6189..7094)
    /note="ORF_f301"
    /codon_start=1
    /label_table=11
    /protein_id="AAAB3042.1"
    /db_xref="GI:887811"
    /translation="MDSARALJARGWGVSLVSRCLVSRSAQLHVILRTDDMMGRRS
    RHTDDVLRIRHVIYELPTGYRVALLRROAELDGMALINAKRYRIRRONALL
    LERKPAVPSKRAHGRVAVKESNORWCGDFECDDNERLRYFPLALCCDREALHW
    AYTGGFNSERTYQDVMGAVERRFGNDLSSPVEWITDNGSCYRANETQFARMGLLE
    PNATVRSPESSNGIASEFVKTIKRDITSIKPDGLTAAKNLALEAEHNEWHPHAL
    GRSRPREYLRQACNGLSNDKLEI"
    complement(7052..7462)
    /note="ORF_f136"
    /codon_start=1
    /label_table=11
    /protein_id="AAAB3043.1"
    /db_xref="GI:887812"
    /translation="MIVLILFRLVIGEQMIDVLGPEKRRRTQEKIAIVOOSEFPG
    MIVSLVAROHGVAASQLFLMRKOYOGSGLTAVAGQOVYPASLAAAMQIKELQRL
    GKTMENELKRAVEYGRAKKNIHAAPLPGGE"
    complement(7506..7805)
    /note="ORF_f99"
    /codon_start=1
    /label_table=11
    /protein_id="AAAB3044.1"
    /db_xref="GI:887813"
    /translation="MTSGLERLSNLSKQDSVFSVDLREAKVNEIDETLSTRLNHL
    IDKGERITLQDLGSGSPGLEKDGHYREADDALINVIYPTLSTKINRRKEQV"
    complement(7851..8687)
    /note="was f268 and f157p before splice"
    /codon_start=1
    /label_table=11
    /protein_id="AAAB3045.1"
    /db_xref="GI:887814"
    /translation="WVIAFNKRIMYEMKKSALITLLAGTISGCTCOATGAETSE
    STPNLINSIGAATDKSYLSLNDPKYPTNYLHSAKLQNEIKSHITKDIQGLITTE
    NTRKLTLEKPMWGTFLASTFBDKTAATHYDAVLRSLDGLWYMLVSDQGSVAK
    KYLLTMDWSPDQIKRMQDVISNPKRLDGIPOVNAVAHIRFDSNSPYMADVOEGK
    POLMNHKQNDALGLYLDLLIQAINSTGINAEDMQGDRKLSVALITAYLDKANFYME
    DS"
    complement(8664..8962)
    /note="corresponds to X52798 (ECTRNAG); 4 sequence
    differences outside tRNA"
    complement(8801..8874)
    /gene="giyu"
    /complement(8801..8874)
    /gene="giyu"
    /product="tRNA-gly"
    /note="codon recognized: GGG; aa: Glycine; CG Site No.
    672; alternate gene names su36, sufd, suna, sunb, supt"
    complement(8953..9732)

```

```

    /note="UUG start"
    /codon_start=1
    /label_table=11
    /protein_id="AAAB3046.1"
    /db_xref="GI:887815"

Alignment Scores:
Pred. No.: 0.000523 Length: 55175
Score: 238.00 Matches: 78
Percent Similarity: 39.628 Conservative: 46
Best Local Similarity: 24.928 Mismatches: 103
Query Match: 14.258 Indels: 86
DB: 1 Gaps: 8

US-10-018-706-2 (1-322) x ECU28375 (1-55175)
QY 16 LysArgLeuGlyLeuIlePheGlyValIleThrCysIleLeuAlaGlyCysAlaSer 35
    |||||.....:
Db 9684 AATCTCTGGGTATCGTATGTTGATCGGTGACGCTGTTTGGCGGGCTGT----- 9631
QY 36 LysProThrTyrAsnSerThrSerGlySerGlySerHisArgThrSerGlyGly 55
    |||||.....-TCGGGTACCAATCATCCGATCAGAGAAAGTATCCGCG
Db 9630 -----
QY 56 LeuAlaIleGlySerGlnValIleThrAspSerGlnGlyValProAsnArgTyrGlnVal 75
    .....-TCGGTTTACCGGTG
QY 76 LysGlnGlyAspThrValSerIleAlaGlnArgTyrGlyLeuAsnTrpArgGluIle 95
    |||||.....:
Db 9576 AAACGGGGGATACGCTATATGTTATTCGGCACACCGGAAACCGTAAAGAACTG 9517
QY 96 GlyHisIleAsnAsnLeuAsnSerSerTyrThrIleTyrThrGlyGlnTrpLeu 115
    .....:
Db 9516 GCGCGACTGACGCGCATTTCCCGCCCTTACACCATGGAAGTTGGTCAAGAACTAAACTG 9457
QY 116 TrpSerGlyAspLeuIleYs-----ValArgIleAspSerIleSerSerGlyVal 131
    .....:
Db 9456 -----GTTGGGGCGCAAGAAAGTACAGTATTAACCGTAATCAACCCCAATTAAGACACC 9403
QY 132 AsnThrAlaHis-----ThrProSerProValAlaValGlnSerSerArgProProValGln 150
    .....:
Db 9402 AAACCGCATCGGTTATCACCGTATCATCGAGCGGTACGAAATCATCTGGCGCGCATGAGG 9343
QY 151 GlnHisProAlaValGlnIleYsProThrProProValAlaValAlaValYsYsProThrPro 170
    .....:
Db 9342 CAACGTTGTTGGTTATGCGCAACGACAGGAGAAAGTTATCATG----- 9301
QY 171 ThrProProValValGlnGlnProAlaProValAlaProProValThrGlnAlaProPhe 190
    .....:
Db 9300 -----CGGTAT 9295
QY 191 AlaThrGlySerSerGlyValMetGlnPheArgTyrProValGlyAlaThrAsnProVal 210
    .....:
Db 9294 TCGACACGACAGAGCGCGC----- 9277
QY 211 ValArgArgPheGlyThrAlaThrValAlaGlySerThrValThrSerAsnGlyMetTrp 230
    .....-ATAAAGGAGATTGAT 9262
Db 9276 -----
QY 231 PheSerGlyArgAspGlyAspLeuIleAsnAlaSerAsnAlaGlyThrValIleGlnAla 250
    |||||.....:
Db 9261 ATCTCAGCTCCACGGGGGTACACTATTTACCGCGCGGTGAGAAAGGTGATGTG 9202
QY 251 AspHisAsnMetAspGlyAlaSer-----IleValIleGlnHisThrAsnGlyPheVal 268
    .....:
Db 9201 GCGAACCACTGGGTGATACGCTATCATCATCATTAATTAACACAGTGAAGTTACATT 9142
QY 269 SerSerTyrIleHisIleLeuYsAspAlaGlnValYsThrGlyAspThrValArgThrGly 288
    .....:
Db 9141 ACGGCTTACGCCCATATATGACACGATCGTGTAAATTAATGCGGCAAGCGTGAAGCGTGGG 9082

```


CDS
3057..3905
/gene="EC3654"
/note="similar to YOCB ECOLI g1|1789158 percent identity
98 in 282 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB37077.1"
/db_xref="GI:13363126"
/translation="MSSYANHQAAGITLCKSDYRDYDTPASILQCVPRSLNDPGL
KADNPFGQDTWTLTLESLMNAKGLPVAVGVHVDLYTSLNLESKFLYINSEFQ
TFRNNDVEVQTLERDLSTCAQGVSAVALRDLERDQPSIGHNGTCIDQDDITINQ
EFTFDLENATSGEKVETLVSHLKSNCILITHQPMWSIQIQYGRQIDREKLTAY
LVSFRHNEHEQCVERIFNDILRQCPKLSVARYTRSGDINDPWSNDFVST
TRLVRQ"
gene
4017..5381
/gene="EC3655"
4017..5381
/gene="EC3655"
/note="similar to YGDH ECOLI g1|1789159 percent identity
99 in 454 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB37078.1"
/db_xref="GI:13363127"
/translation="MTHTISPLGSMMDLSQLEVDMLKRTASDLYQIFNCSIALVNS
GSLTNSKELSLRFENFDINVLRRERGVKELINPEEAVDRRIILADLANFAVLR
DILFVQIHNTRFPLNDNSVHTNLVFSILRNRAVLHVSAPVMVVCWGSIN
ENEYLARVGNQGLRELNICTGCGGAEAPMKGAAGVHAQORVDSRFICTEBS
IIAEPNPPLVNEILIMPDIKRLBAFVRIAHGIIIPGGVTAABELLYLGLIMNA
MKQVPLITGPKESADYPRVLDPEVHTLGENARHRYIITDDAEVAVQKMSRP
LYKENRDYGDAYSFNMSRIAPDLQVPEFSHNNANLKLYPDQVEVLAADLRRA
SGIVAGNVKEVGIRALIEFGPKINGDKEMRMDDLQGFVAQHRMKLPGSAVYIPCY
EICT"
gene
5938..7227
/gene="EC3656"
5938..7227
/gene="EC3656"
/note="similar to SDAC ECOLI g1|1789160 percent identity
100 in 429 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="probable serine transporter"
/protein_id="BAB37079.1"
/db_xref="GI:13363128"
/translation="METTOSTIASKDSRSAMRKTDDMMMLGYTAIGAGVFLPIN
AGVGMIPILIMAILAPMFPFAHGRITRPVLSKNGDGEITVEHEHFGIGKLT
LYPFAIYPLILVSVAITTVTSFMSHQGMPPRAIISLILVGMETIVYRGEOM
IVKMSILVPEFVGLMLALYLIPQNGALSTLSDTASATGNGIMTMLALIPVM
VPSFMSPIISSPAVAKREYGMARCKKILAFHIMVLTVMFVPSCVISLTPA
DLAAKEONISILSYLANHFAVIMAMAILIATITKFLGHYIGARSGFNGMYK
SLRGKSEIKNLRIITLFLMTVITVITLNPISLGMETIGPPIANILFLMPVY
ALOKVPMKRYSHISNVFVVMGLAISIFSLRS"
gene
7285..8652
/gene="EC3657"
7285..8652
/gene="EC3657"
/note="similar to SDAA ECOLI g1|1789116 percent identity
77 in 455 aa, similar to SDAB ECOLI g1|1789161 percent
identity 99 in 455 aa (Conserved in E.coli K-12)"
Alignment Scores:
Pred. No.: 0.00298 Length: 266658
Score: 238.00 Matches: 78
Percent Similarity: 39.62% Conservative: 46
Best Local Similarity: 24.92% Mismatches: 103
Query Match: 14.25% Indels: 86
DB: 1 Gaps: 8

US-10-018-706-2 (1-322) x AP002563 (1-266658)
QY 16 LysArgGluGlyLeuIlePheGlyValIleThrThrCysIleLeuAlaGlyCysAlaSer 35
DB 93506 AAATCTGGGATACGATGATGTTATTCGGTTCGACTGCTTTGGCGGCTGT----- 93453
QY 36 LysProThrTyraSerSerThrSerGlySerHisArgThrSerGlySerGly 55
DB 93452 -----TCGGGTACGAATCATCCGATACAGGAAGTATTCGGGC 93414
QY 56 LeuAlaIleGlySerGlnValIleThrAspSerGlnValProAsnArgTyGlnVal 75
DB 93413 -----TCCGTTTACACCGCTG 93399
QY 76 LysGlnIleAspThrValSerValIleAlaGlnArgTyGlyLeuAsnTrpArgGlnIle 95
DB 93398 AAACGGGGGATACGCTATATCGATTTCGGACCAACGGGAACACGTAAGAGCTG 93339
QY 96 GlyHisIleAsnAsnLeuAsnSerSerTyThrIleTyThrGlyGlnTrpLeuThrLeu 115
DB 93338 GCGGACGTGAACGGCATTCCTCCCTTACACATTGAAGTGTGAGAAACTAAACTG 93279
QY 116 TrpSerGlyAspLeuLys-----ValArgGluArgSerIleSerSerGlyVal 131
DB 93278 -----GTCGGCGCGAAAGTAGACAGTAGACAGTAATCAACCGCAATCAACGACC 93225
QY 132 AsnThrAlaHis---ThrProSerProValAlaValGlnSerSerArgProProValGln 150
DB 93224 AAAACCGCATCGCTTACACCGCTATCAGCGGTACCGGAATCTTCGCGCGCAGTAGGG 93165
QY 151 GlnHisProAlaValGlnLysProThrProProValValValLysProThrPro 170
DB 93164 CAACGTTGTTGTTATGCGCAACGACGAGGAAGTTATCATG----- 93123
QY 171 ThrProProValValGlnGlnProAlaProValAlaProProValThrGlnAlaProPhe 190
DB 93122 -----CCGTAT 93117
QY 191 AlaThrGlySerSerGlyValMetGlnPheArgTyProValGlyAlaThrAsnProVal 210
DB 93116 TCACACGACGATGCGCGC----- 93099
QY 211 ValArgAspPheGlyThrAlaThrValAlaGlySerThrValThrSerAsnGlyMetTrp 230
DB 93098 -----AATAAAGGATTCAT 93084
QY 231 PheSerGlyArgAspGlyAspLeuIleAsnAlaSerAsnAlaGlyThrValIleGlnAla 250
DB 93083 ATCTCAGCTCCACGGGGGTACACCTATTCGCGCGGTGCAGGAAGTGTGTATGTG 93024
QY 251 AspHisAsnMetAspGlyAlaSer-----IleValIleGlnHisThrAsnGlyPheVal 268
DB 93023 GGCACACGAGTGGTGGCTACCGGTATCTCATGTATGAATGAACACAGTGAAGATTACATT 92964
QY 269 SerSerTyThrIleHisIleLysAspAlaGlnValLysThrGlyAspThrLysArgTyGly 288
DB 92963 ACAGCTTACGCCCTAATGACAGATGCTGTAAATATGAGCAACGCTGAAGCTGGG 92904
QY 289 GlnArgIleAlaSerMetLysAsnGlnProSerGlyValAlaLeuPheGlnPheArgIle 308
DB 92903 CAATAAATCGCACCATGAGGGGACAGAGGAGGCGCATTCGCTTCGATTCACAGATT 92844
QY 309 SerArgAsnGlyValTyValAspProLeuThrValLeu 321
DB 92843 CGTTACCGTCAACGCGCAATGATCGCTACGTTACTTG 92805
RESULT 15
AE012272 10773 bp DNA linear BCT 23-MAY-2002
LOCUS Xanthomonas campestris pv. campestris str. ATCC 33913, section 180
DEFINITION of 460 of the complete genome.
ACCESSION AE012272 AE008922
VERSION AE012272.1 GI:21112795

/codon_start=1
 /transl_table=11
 /product="lipoprotein"
 /protein_id="AAM41003.1"
 /db_xref="GI:21112801"
 /translation="MEGNEBMRSAOMKIVRTAVWLVAAGLITACSSATVYRNSNSS
 /GGASRPSAPSPVPRPGATVYVGRGTIVAISRNTITPDLAANGKSSPTPIYV
 /GOTLKYPGASKPGASAPVAGCTVYPRRPAPIAAVSSGFSRMPADQVVGTVT
 /TSETTRQGVADIGASGVARRAADGVVYVSGLVGELIITKHNDQMLSAYGHNRK
 /RLNLRGQSVKQVQIAEWGRSGAARDMLFEIRYNGKVPDPLVLPKK"
 complement(5111..5485)
 /gene="XCC1710"
 complement(5111..5485)
 /gene="XCC1710"
 complement(5111..5485)
 /note="identified by sequence similarity; putative; ORF
 located using Blastx/Glimmer/Genemark"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AAM41004.1"
 /db_xref="GI:21112802"
 /translation="MPLNQEHDPDYTALRAADGRHAKVNEQILOOSFILMPDELVEHM
 /PVSYGOQIPRRMDVALNPAYVILGGERQOFPSTVLAACITRIGLEAMTNAAA
 /ARTVNVLASEGRRVLAIVGG"
 complement(5502..5807)
 /gene="XCC1711"
 complement(5502..5807)
 /note="identified by sequence similarity; putative; ORF
 located using Blastx/Glimmer/Genemark"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AAM41005.1"
 /db_xref="GI:21112803"
 /translation="MSVITSNONFRFLGQAHDLKALLQNGGKVTAPPAVEEDVE
 /RHELKVVAADRETRDALIAIEIHTGICALVQRIGHVALIYRPSKROIVLPGRG"
 5867..6505
 /gene="ftsJ"
 /note="XCC1712"
 5867..6505
 /gene="ftsJ"
 /note="identified by sequence similarity; putative; ORF
 located using Blastx/Glimmer/Genemark"
 /codon_start=1
 /transl_table=11
 /product="cell division protein"
 /protein_id="AAM41006.1"
 /db_xref="GI:21112804"

Alignment Scores:	
Pred. No.:	0.000139
Score:	234.50
Percent Similarity:	40.73%
Best Local Similarity:	27.48%
Query Match:	14.04%
DB:	1
US-10-018-706-2 (1-322)	x AEO12272 (1-10773)
	Length: 10773
	Matches: 83
	Conservative: 40
	Mismatches: 108
	Indels: 71
	Gaps: 11

QY	25	llelthrhtrcyvllleuallaglycyalaaserlyspfothrttyaasmerthsergly	44
		:::	
Db	4224	ctgacgccttcacacgacgcacagctgctgctgccc-----aaactcggcagctgt	4277
QY	45	serglyserthlsargthrserylseryglyleuallaile-----glysergin	61
		:::	
Db	4278	gctggcccttgcggccacagtcgccgccgcctcctgcgtcccgctccgctgcacac	4333
QY	62	vallelthraespserglnlyvalprobsnarqtyrginallyssglnlyaspthval	81
		:::	
Db	4338	gtg-----aacgtgcacgcggcgacacgcgtg	4366
QY	82	serlysvllealaglnarqtyrglyleuabentprapgtgllleghlhisilleabnleu	101

Db	4365	TATGGCATTTGGCCCGCACAATATACCGCGCCGACCTGGCTGTGAACGGGCTG	4424
Qy	102	AanSerSerTyrrhrlleTyrrhglValnTrpLeuThrleuTrpSerGlyAspLeuLys	121
Db	4425	TCTTACCGCAACAGACTTACCCGGGGAGACGCTCAAGCTGTATCCG-----	4472
Qy	122	ValArgIuarSerIleSerSerGlyValAsnThrAlaHisTrpProSerProValAla	141
Db	4472	-----	4472
Qy	142	ValGInSerSerArgProProValGInGInHisProAlaValGInLysProThrProPro	161
Db	4473	-----CCGGCGCGACCAAGCTGGTGCAAGT	4499
Qy	162	ValValValValLysLysProThrProThrProProProValGInGInProAlaProval	181
Db	4500	GCACCGGTGGCGCAATGGCGGACCGGTGTCGCGC-----CGCCGGCGCGACCGATT	4553
Qy	182	AlaProProValTrhGluAlaProPheAlaThrGlySerSerGlyValMetGlnPheArg	201
Db	4554	CGGGCGCCGGTGC-----AGTAGCGGC--TTCAGCTGGGCG	4586
Qy	202	TyrTrpValGlyAlaThrAsnProValValArgArgPheGlyThrAlaThrValAlaGly	221
Db	4587	TGGCGC-----GCCGACGGCGGTGTGGTGAGACCTTC-----GTCAACCGGC	4628
Qy	222	SerThrValThrSerAsnGlyMetTrpPheSerGlyArgAspGlyAspLeuIleAsnAla	241
Db	4629	GAGACC--ACCAAGCAAGGCGGTGACATTTCGCGGTGCCAGCGGCAAGCCGTGGCGCC	4685
Qy	242	SerAsnAlaGlyThrValIleGlnAlaAspHisAsnMetAspGlyAlaSer-----Ile	259
Db	4686	GCCGCCGACGGCGCTGGTGCTATTCGGCGCGCGCGCTGTGCGGCTACGGGAGACTGATC	4745
Qy	260	ValIleGlnHisThrAsnGlyPheValSerSerTyrrIleHisIleLysAspAlaGlnVal	279
Db	4746	ATCATCAAGCAACAGATCAAGTGGCTGGCGCTACAGCCACCAACCGCAACGCGCTGCTC	4805
Qy	280	LysThrGlyAspThrValArgTrhGlyGlnArgIleAlaSerMetLysAsnGlnProSer	299
Db	4806	AACCAAGGCCAGACGCTGAAGCCGGCGCAGAGATCCCGAATGGGCGCGACGGCGGCC	4865
Qy	300	GlyAlaIalaLeuPheGlnPheArgIleSerArgAsnGlyValTyrrValAspProLeuThr	319
Db	4866	GCCCCGCAATGCTGCACTTGAATCCGCTACACAGCGCAAGCCGGTGAACCCGCTGCTG	4925
Qy	320	ValLeu 321	
Db	4926	TATTTG 4931	

RESULT 16	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	MEDLINE
AE004905/c	AE004905	10822 bp DNA linear BCT 30-AUG-2000	AE004905	AE004905	GI:9951195	<i>Pseudomonas aeruginosa</i> PA01, section 466 of the complete genome.	<i>Pseudomonas aeruginosa</i> . <i>Pseudomonas aeruginosa</i> Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae; Pseudomonas.	1 (bases 1 to 10822)	Stover, C.K., Pham, X.O., Erwin, A.L., Mitsuoguchi, S.D., Warrenner, P., Hickey, M.J., Brinkman, F.S., Huftagel, W.O., Kowalik, D.J.J., Lagrou, M., Garber, R.L., Goltry, L., Tolentino, E., Westbrock-Nadman, S., Yuan, Y., Brody, L.L., Coulter, S.N., Folger, K.R., Kas, A., Larbig, K., Lim, R., Smith, K., Spencer, D., Wong, G.K., Wu, Z., Wu, Z., and Paulsen, I.T.	Complete genome sequence of <i>Pseudomonas aeruginosa</i> PA01, an opportunistic pathogen	Nature 406 (6795), 959-964 (2000)	20437737

FEATURES
source
1. 10822
/organism="Pseudomonas aeruginosa"
/strain="PAO1"
/db_xref="taxon:287"
86. .745
/gene="PA4918"
86. .745
/gene="PA4918"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAG08303.1"
/db_xref="GI:9951196"
/translation="MNRPSKIASPDVPAOKGFTPLCPDELPEVGDAIGALNQLASR
AALRIGSKDAHSPKAAWVRSAROMLRLPALDITWVSHCVGPTGPGFIDELP
IDYDPFVKGVEPDHIEHYGACYHDLARSTGVTEYKONGSVHVLVIGALDYCVN
TAIQLRAGSEVLLYLPACRAIADPTQRACDNRDAGVILCADLRELDALAGADSE
K"
748. .1947
/gene="pncB1"
/note="PA4919"
748. .1947
/gene="pncB1"
/codon_start=1
/transl_table=1
/product="nicotinate phosphoribosyltransferase"
/protein_id="AAG08304.1"
/db_xref="GI:9951197"
/translation="MAESVPAERIVQNLDTDFYKLTMMQAVLHNPNAEWEFERCR
NADLRPYLAIRYQLEIRLAIEVTLADQLAFLEKIPMKDPIRFLSIFERNLRYHT
GIEGQALIRLGRFVHILFEVPLAIVSEVRNRYREVEVLEQVEQLYRKIDMS
AQSSEELAEFOVADFGTRRRFSYRTOEYVHILKRDPPGFRVGTSNVHALREYDLK
IGMAHEMLMAHQQLGRLVDSQQALDCWREYRGQIGIATDCITMDAFLDFDLG
FALFDGLRHDSDGPLMAEKALNHYRLGIDPLSKTLVPSDGLMDKALQLEFALG
KINVSFGIGNLFTCDIPGVPEPMNIVLKWTACNGHPAKISDAPGKTCCRDENFVAYLR
HVEVPA"
1976. .2803
/gene="nabB"
/note="PA4920"
1976. .2803
/gene="nabB"
/codon_start=1
/transl_table=1
/product="NAD dependent NAD synthetase"
/protein_id="AAG08305.1"
/db_xref="GI:9951198"
/translation="WQOIQRIQALQVQPPQSEADVQACIARIATFOOCLKDSGL
KTVLVGSGGVDSLTGILAQRAVEQLREQDQAYFPIARLPYQVQDADQASL
ATIRADEQVTNIGSPVKALAEQLALEGLPEPASPDSYFVIGIKIRIMVQYALAGAR
GLVIGTDHAAAEVGFTEKGDGACDLAIGSLGAKQVRLARALGAPENLVKIKPT
ADLEDLRPGHDEASHGVYAEIDAFLHGOPLREBAARVIVDTYHKTQHKRELKPA"
2937. .3860
/gene="PA4921"
2937. .3860
/gene="PA4921"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AAG08306.1"

gene
CDS
/db_xref="GI:9951199"
/translation="MRTRLPALLGLVLLAGACGHTSPLLAPVRQIHAFGDSYDNCE
SORLTREMLAKGIAGQALPGEVYQWQNSGPTAVEVLARQIAGLADHAAVAGAKG
ANRYGQMAVYHHTGAGQVADVYLAITLDGKPVDAQHPIFVSANDFEHEDFAGEOP
LBIQAGSSVANIIRAAVORGEAGARRFLVSSSTDLSVPAAVAGNRVERAQRYLQAVN
ASLPLOALIRKTRGELSMFPHLTFPSRLRNPARYGVIELDAPCFQPTQPSVRPACA
NDOXYFMEWHMPTRRVHQLAGEAAMARFAR"
complement (3929. .4375)
/gene="azu"
/note="PA4922"
complement (3929. .4375)
/gene="azu"
/codon_start=1
/transl_table=1
/product="azurin precursor"
/protein_id="AAG08307.1"
/db_xref="GI:9951200"
/translation="MLRKAAVSLSLASAPLLAECSVDIQGDMQFNINATITDK
SCQFTVNLSHRGNLPKRWGHWVLSIADQGVVTDMAAGLDKDYLPDPSRVIA
HKLIGSGKDSVTFDVSFKKEGQYMFCTPFGHSLMKLTLLK"
4651. .5238
/gene="PA4923"
4651. .5238
/gene="PA4923"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AAG08308.1"
/db_xref="GI:9951201"
/translation="MTLRSVCEFCGSPGASPVYQEAVALGRHLAERGLTVYGGGA
VGMGVADAAALAAAGEVIGITPOSLOEAEIIGKGLTRLEVVDGMAARLAAEELDA
FLAPGLGTLRELFQVMTWGOGLYAKPLGLELVGFPDPLITFLDHLVDERFVAB
HRCMLORGASPEALDALAAMPVSVAPKWVDRPQ"
complement (5251. .5946)
/gene="PA4924"
complement (5251. .5946)
/gene="PA4924"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AAG08309.1"
/db_xref="GI:9951202"
/translation="MFRLLILVLLAVAGCSKAPGASVYKRYRGERVYRGDTL
YSIATRHGMVQKOLARANGIRPPYAKVQVAFEDKSTVYASSSSSTRARKPP
GPEVATVIRIGWQMPKGPVIRRESSSKLNKGIRIAGTLQPVQASLAGKVFAVNNKR
GYNVATVIRIGWQMPKGPVIRRESSSKLNKGIRIAGTLQPVQASLAGKVFAVNNKR
GRLDPLSLIPPS"
complement (6133. .6984)
/gene="PA4925"
complement (6133. .6984)
/gene="PA4925"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AAG08310.1"
/db_xref="GI:9951203"
/translation="MDDTIVNVTGKINALMTVSOYAFGKILIAALAFWVIGRW
LIGFAVGMQKSLERQDPTVYRVGSFTVLNVLVGLVGVGKFGIQTSLAALIA
AVGLAIGMMSGLLANLAAGGFTIVLRPFVQGFIVAGGTGKVEKGLTSLAALIA
NVLTWVGNNKIRGDNIONFHPNFRVELEKQALSGADYOALATALKORVAAALPNVIG
EPVVDYELIEFNLVGCVLAVRPYCHNDHWQVYFDINKYIKDALGADFAPMPAQYVI
VOOSAAQ"
complement (7235. .8170)
/gene="PA4926"
complement (7235. .8170)
/gene="PA4926"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AAG08311.1"
/db_xref="GI:9951204"
/translation="MSAPALRSVPLQGARVYVLAHDTHTRYRYPVLSISROLAHMREPC
AMQCHERRALLIEGSPSRREDGHDAFGNPLSRILAFRPHALRVSARLQVEVPAHAA

gene
 CDS
 complement (8170..10662)
 /gene="PA927"
 /complement (8170..10662)
 /gene="PA927"
 /codon_start=1
 /transl_table=1
 /product="conserved hypothetical protein"
 /protein_id="PA608312.1"
 /db_xref="GI:9951205"

OTASPMETVAAGLAFSGALSAELACRYVESYVYIQRFAEFAADCEPPGR
 LVAALAMERKIFREFAADQATQVADPLEVERRGVCCDFAMILACIARGLA
 RYVGYLLTPRPOARLIGADASHAVVSYCPENGVDPDPTNLLPDEHITLWAG
 RDSVSLRGLGSSHDPEVRYTLPIDETAG"
 /transl_table=1
 /product="conserved hypothetical protein"
 /protein_id="PA608312.1"
 /db_xref="GI:9951205"

BASE COUNT 1843 a 3728 c 3606 g 1645 t
 ORIGIN

Alignment Scores:
 Pred. No.: 0.0015 Length: 10822
 Score: 234.00 Matches: 68
 Percent Similarity: 41.18% Conservative: 37
 Best Local Similarity: 26.67% Mismatches: 90
 Query Match: 14.01% Indels: 60
 Gaps: 6

US-10-018-706-2 (1-322) x AE004905 (1-10822)

QY 69 ValProaanhgryglnvalynglnlyasphthvalserlyleaglinalgrytyr 88
 Db 5853 GTGGCGCGGCGGATCGCTCAAGCGCGGCGACCCCTCTATTCGATCGACCCCGGAC 5794
 QY 89 Glyleuasentpraggluilegylhisleasnasleuasensersertyrthiletyr 108
 Db 5793 GGCTGGAACCTACAGACCTGGCGCGGCGGCGACCTCCGCGCTACGCGGTCAAG 5734
 QY 109 Thrglyglntrpleuthrleutprserglyaspleuysvalargluarserileser 128
 Db 5733 GTGGCGCGGCGGATCGCTCAAGCGCGGCGACCCCTCTATTCGATCGACCCCGGAC 5677
 QY 129 Serglyvalanthralahsthrproserprovalalavalginsersearaypro 148
 Db 5676 TCCTCGCTACACCCGT----- 5659
 QY 149 Valglnghisproalavalglnlyeprothrprouvalvalvallylyep 168
 Db 5658 -----GCCGCGCAACCG 5647
 QY 169 Thrprothrprouvalvalglnghlnproalaprovalalaprovalalthgluala 188
 Db 5646 CCGGCTCGCGCGCGCTGGTAC----- 5623
 QY 189 Probhealathrglyserseglyvalmetglnpheargtyrprovalglalyalathasn 208
 Db 5622 -----CTGGCGCGCTGGCGAGTGGCGGATG-----AAGGG 5593
 QY 209 Provalvalargargphneglythralathralaglyserthvalthserasnly 228
 Db 5592 CCGGTATCGCGCGCTTC-----TCAGTTCGCAAGCTCAACAGGCG 5548

QY 229 Mettrpbeserglyargasglyaspleuileasnalaserasnalaglythvalille 248
 Db 5547 ATCCGATCCCGGACCGCTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5488
 QY 249 GlnlAasphlsamnetasglyalaser-----llevalleglnhsthrasnly 266
 Db 5487 TTGCGGCTGAACACACACCTGCTACCGGACCTGTGATCATCCGACCGGATCTCG 5428
 QY 267 Phevalsersegyrthlehisilelyasplalaglnvallysthrghlyaspthvalarg 286
 Db 5427 TACACGACACCTACGCGGACACACGCGGCTGCTGTGAGAGAGGCGGATGTGCGC 5368
 QY 287 Thrglyglnarglleaasermetlyasnglnproserglyalalaleupheglnphe 306
 Db 5367 AAGGCGGAGAAATCGCGGCGGCTTCTCGGATCCGACCGGCGGTGACGTCTACTTC 5308
 QY 307 Argllsearayasnlyvaltyrvalasproleuthrvalleu 321
 Db 5307 GAGATCGGCGGACGCGGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGG 5263

RESULT 17

AE008839/c 22400 bp DNA linear BCT 31-JUL-2002

LOCUS

AE008839 AE006468

DEFINITION

Salmonella typhimurium LT2, section 143 of 220 of the complete genome.

ACCESSION

AE008839.1 GI:16421569

VERSION

AE008839.1

KEYWORDS

Salmonella typhimurium LT2.

SOURCE

Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Salmonella.

ORGANISM

Salmonella typhimurium LT2.

REFERENCE

1 (bases 1 to 22400)
 2 (bases 1 to 22400)

REFERENCE

McClennan, M., Sanderson, K.E., Spieth, J., Clifton, S.W.,
 Letellier, P., Courtney, L., Porwollik, S., All, J., Dancie, M., Du, F.,
 Hou, S., Layman, D., Leonard, S., Nguyen, C., Scott, K., Holmes, A.,
 Grewal, N., Mulvaney, E., Ryan, E., Sun, H., Flores, L., Miller, W.,
 Stoneking, T., Nhan, M., Waterston, R., and Wilson, R.K.
 Complete genome sequence of Salmonella enterica serovar Typhimurium
 LT2
 Nature 413 (6858), 852-856 (2001)

TITLE

Salmonella typhimurium LT2.

REFERENCE

1 (bases 1 to 22400)
 2 (bases 1 to 22400)

REFERENCE

McClennan, M., Sanderson, K.E., Spieth, J., Clifton, S.W.,
 Letellier, P., Courtney, L., Porwollik, S., All, J., Dancie, M., Du, F.,
 Hou, S., Layman, D., Leonard, S., Nguyen, C., Scott, K., Holmes, A.,
 Grewal, N., Mulvaney, E., Ryan, E., Sun, H., Flores, L., Miller, W.,
 Stoneking, T., Nhan, M., Waterston, R., and Wilson, R.K.
 Complete genome sequence of Salmonella enterica serovar Typhimurium
 LT2
 Nature 413 (6858), 852-856 (2001)

TITLE

Salmonella typhimurium LT2.

REFERENCE

1 (bases 1 to 22400)
 2 (bases 1 to 22400)

REFERENCE

McClennan, M., Sanderson, K.E., Spieth, J., Clifton, S.W.,
 Letellier, P., Courtney, L., Porwollik, S., All, J., Dancie, M., Du, F.,
 Hou, S., Layman, D., Leonard, S., Nguyen, C., Scott, K., Holmes, A.,
 Grewal, N., Mulvaney, E., Ryan, E., Sun, H., Flores, L., Miller, W.,
 Stoneking, T., Nhan, M., Waterston, R., and Wilson, R.K.
 Complete genome sequence of Salmonella enterica serovar Typhimurium
 LT2
 Nature 413 (6858), 852-856 (2001)

TITLE

Salmonella typhimurium LT2.

REFERENCE

1 (bases 1 to 22400)
 2 (bases 1 to 22400)

REFERENCE

McClennan, M., Sanderson, K.E., Spieth, J., Clifton, S.W.,
 Letellier, P., Courtney, L., Porwollik, S., All, J., Dancie, M., Du, F.,
 Hou, S., Layman, D., Leonard, S., Nguyen, C., Scott, K., Holmes, A.,
 Grewal, N., Mulvaney, E., Ryan, E., Sun, H., Flores, L., Miller, W.,
 Stoneking, T., Nhan, M., Waterston, R., and Wilson, R.K.
 Complete genome sequence of Salmonella enterica serovar Typhimurium
 LT2
 Nature 413 (6858), 852-856 (2001)

FEATURES

Location/Qualifiers

Coding sequences below are predicted from manually evaluated
 computer analysis, using similarity information and the programs;
 GLIMMER; <http://www.tigr.org/softlab/glimmer/glimmer.html> and
 Genemark; <http://opal.biology.gatech.edu/Genemark/>
 EC numbers were kindly provided by Junko Yabuzaki and the Kyoto
 Encyclopedia of Genes and Genomes; <http://www.genome.ad.jp/kegg/>,
 and Pedro Romero and Peter Karp at Ecocyc;
<http://ecocyc.org/ecocyc/>

The analyses of ribosome binding sites and promoter binding sites
 were kindly provided by Heladia Salgado, Julio Collado-Vides and
 ReguonDB;
http://kinich.cifn.unam.mx:8850/db/regulondb_intro.frameset

This sequence was finished as follows unless otherwise noted: all
 regions were double stranded, sequenced with an alternate
 chemistries or covered by high quality data (i.e., phred quality >=
 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by sequence
 from more than one m13 subclone.

gene	complement (7237. .8320)
CDS	/gene="STM3026" complement (7237. .8313)

Alignment Scores:	
Pred. No.:	0.00041
Score:	232.50
Percent Similarity:	40.19%
Best Local Similarity:	24.04%
Query Match:	13.82%
DB:	1
Length:	22400
Matches:	76
Conservative:	49
Mismatches:	99
Indels:	87
Gaps:	9

US-10-018-706-2 (1-322) x AE008839 (1-22400)

QY	18	LeuGIleuIlePheGlyValIleThrThyValIleLeuAlaGlyCysAlaSerIlePro	37
DB	18880	CTGGGATTTCGGGCTGCTTATGCACAGGGCTGCTGGCGGGGT	18833
QY	38	ThrTyrAsnSerThrSerGlySerGlySerHisAlaThrSerGlySerGlyLeuAla	57
DB	18832	-----TCCAGTAATTCGGGCTCAGAGACGAT-----TCCGGCTCC-----	18797
QY	58	IleGlySerGlnValIleThrAspSerGlnGlyValProAsnArgTyrGlnValIleGln	77
DB	18796	-----GTCATACCGTTAAGCC	18779
QY	78	GlyAspThrValSerIleValIleAlaGlnArgTyrGlyLeuAsnTyrArgGlnIleGlnHis	97
DB	18778	GGGATATCGTTGATATGATATGATGCGCGGCTACGGGAACACGCGTGAAGAACCTGGCCGA	18719
QY	98	IleAsnAsnLeuAsnSerSerTyrThrIleTyrThrGlyGlnTyrLeuThrLeuTyrSer	117
DB	18718	CTGAATTCACATTTCTCCGCGCTTATGATTTAGAGTGGGACGAGGATTAAGTGC--CGC	18662
QY	118	GlyAspLeuLys-----ValArgGlnArgSerIleSerSerGlyValAsnThr	133
DB	18661	GGCAGCGCCAAAGATGACACTTTCACACCGCAAAACACGACATTAACACACACAAAACCT	18602
QY	134	AlaHisThr--ProSerProValAlaValGlnSerSerArgProProValGlnGlnHis	152
DB	18601	GCTGCGGCTGACACCGTCATCTTGGTGGCCAAATCCTCCTGGCGCGGCTGGGGGACGCT	18452
QY	153	ProAlaValGlnLysProThrProProValValValLysBysProThrProThrPro	172
DB	18541	TGTTGGTATAGCGCTCGTAAACGTAAGTGAATTTA-----	18506
QY	173	ProValValGlnGlnProAlaProValAlaProProValThrGlnAlaProPheAlaThr	192
DB	18505	-----CCTATTCCACC	18494
QY	193	GlySerSerGlyValMetGlnPheArgTyrProValGlyAlaThrAsnProValValArg	212
DB	18493	GCGGAGGCGCGA-----	18482
QY	213	ArgPheGlyThrAlaThrValAlaGlySerThrValThrSerAsnGlyMetTyrPheSer	232
DB	18481	-----AATAAGGACATCGATTGCG	18461
QY	233	GlyArgAspGlyAspLeuIleAsnAlaSerAsnAlaGlyThrValIleGlnAlaAspHis	252
DB	18460	GGGCGCGCGCGAACCGCTTTTATGCGCGGCGGACGCGTAAAGTTGTCATATCGCGGAC	18401
QY	253	AsnMetAspGlyAlaSer-----IleValIleGlnHisThrAsnGlyPheValSerSer	270
DB	18400	CAGCTGCGCGGTTACCGCAATCTCATATCATATTAAGACAAATGAAGATTACATTACGGCT	18341
QY	271	TyrIleHisIleLysAspAlaGlnValLysThrGlyAspThrValArgThrGlyGlnArg	290
DB	18340	TACGCTCATATATGACACCATGCTGTTAATTAAGACAAAGCGTTAAACGGGCGCAAAA	18281
QY	291	IleAlaSerMetLysAsnGlnProSerSerGlyAlaAlaLysPheGlnPheArgIleSerArg	310
DB	18280	ATGCTACACATGGCGACACGATGCGGATCCGTAAGACTACACTTCAGATTCGGTAC	18221

```

Oy      311 AsmglyValIyrValAspProLeuthrValIeu 321
          ::::::::::::::
Db      18220 CGGGCGACGGCCATGCATCCGCTGCATTATCTG 18188

```

RESULT 18	AL62277/c	LOCUS	DEFINITION
AL622777	23050 bp	DNA	linear
			BCT 06-JUN-2002
			<i>Salmonella enterica</i> serovar Typhi (Salmonella typhi) strain CT18, complete chromosome; segment 13/20.

SOURCE ORGANISM	TEST
Salmonella enterica subsp. enterica serovar Typh.	+
Salmonella enterica subsp. enterica serovar Typh	+
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Salmonella	+

REFERENCE
AUTHORS

REFERENCE
AUTHORS

1 (Passes 1 to 230050)
Parkhill, J., Dougan, G., James, K. D., Thomson, N. R., Pickard, D.,
Main, J., Churcher, C., Mungall, K. L., Bentley, S. D., Holden, M. T. G.,
Sebatian, M., Baker, S., Basham, D., Brooks, K., Chillingworth, T. C.,
Comerton, P., Conn, A., Davis, P., Davies, R. M., Dow, L., White, N.,
Barra, J., Fellwell, T., Hamlin, N., Haque, A., Hien, T. T., Holroyd, S.,
Jagels, K., Krogh, A., Larsen, T. S., Leather, S., Mølle, S., O'Gaora, P.,
Perry, C., Quail, M., Rutherford, K., Simmonds, M., Skelton, J.,
Stevens, K., Whitehead, S. and Barrall, B. G.

JOURNAL	Nature 413 (6858), 848-852 (2001)
MEDLINE	21534947
PUBMED	11677608
REFERENCE	2 (bases 1 to 230050)

JOURNAL Submitted (25-OCT-2001) Submitted on behalf of the Salmonella
sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
Hinxton, Cambridge CB10 1SA, UK
COMMENT E-mail: parkhill@sanger.ac.uk

FEATURES

SOURCE

gene

```

309. 1598
/genre="STR3109"
/note="Fasta hit to Q46940 (409 aa), 40% identity in 416
aa overlap
Fasta hit to TDCO_ECOLI (443 aa), 50% identity in 442 aa
overlap
Orthologue of E. coli sdac (SDAC_ECOLI): Fasta hit to
SDAC_ECOLI (429 aa), 95% identity in 429 aa overlap"
/codon_start=1
/transl_table=1
/product="putative serine transporter"
/protein_id="CAD02796.1"
/db_xref="GI:16504050"
/db_xref="SPTREMBL:O8Z435"
/translation="MKTOTSTIASIDRSAMKRTTMMGLGYRIGAGVLPFIN
AGGGMPLIIMALIVPMTFEFAHRLGRLRVLSGKNPGEITVVEHEHPIGAGKIT
LAFPIYPIPLIVYSAIVNTVYESPLTHQALPPRAIISLIVGMNTIVPGEOM
IVAMGILVPPFAVALMLTALYIIPOMNGGALLETISDSASGNGMIMTLMLAIPVM
VFPENSPITSSFAVAKREEVGGAERKSKITLAFPHIMVILVMPFVESCIVLIPA
DLAAKEONISILSYLANHPNAPITIAMAPIIMATITSFLGHYLGARGPQGMWIK
SLRGKSEIINKLAKNTLPHMLVTYITVAITSLTGLTGGPIIANTILMLPMPT
AIOKVPAMKRYSGHISNVFVIMGLAIISAIFSLFS"

```

```

gene      1657..3024
           /gene="STRY3110"
           /note="sdab"
CDS       1657..3024
           /gene="STRY3110"
           /note="Fasta hit to TDCG_ECOLI (454 aa), 73% identity in
           454 aa overlap
           Fasta hit to SDH_ECOLI (454 aa), 77% identity in 455 aa
           overlap
           Orthologue of E. coli sdab (SDH_ECOLI); Fasta hit to
           SDH_ECOLI (455 aa), 94% identity in 455 aa overlap"
           /codon_start=1
           /transl_table=11
           /product="L-serine dehydratase 2 (L-serine deaminase 2)"
           /protein_id="CAD02797.1"
           /db_xref="GI:16504051"
           /db_xref="SPTREMBL:Q82434"
           /translation="MISVFDFIKIGIPSSSHTVGPMKAKGKQFTDDLARHILTDVTR
           VVDVYGLSLTGKHTIDALITMGLAGNLPDVIDALPSFIQDVNTHGRLLANGQ
           HEVEPVDKOMNFADNLSLHENGRIITALAGKVLVSQTYSGIFVDEBFGILT
           NSEPVVVPYKYTAADLQRCQETGLSLGLMMQNEIALHSEKALEOHMFKWEVMS
           GIERGITGTVLGEKLRVPRRAALRMVLVSQDKITDPAVVDWIMFALVNEENA
           AGRVVTAFTNGACGIVPAVLAYDKFIREVNANSLARVMLVASAIGSLYKMAISG
           AEVCGEGVACSMAGIALLGSPAVOCIAAEIAMEHNIQLTCDPVAQGVQVPC
           IERNALASVAVAAARMALRTSEPRVCLDKVLETMYETGKDMNAKRETSRGLAMK
           IVTCD"
gene      3135..3950
           /gene="STRY3111"
           /note="exo"
CDS       3135..3950
           /gene="STRY3111"
           /note="Orthologue of E. coli exo (EX9_ECOLI); Fasta hit to
           EX9_ECOLI (281 aa), 88% identity in 264 aa overlap"
           /codon_start=1
           /transl_table=11
           /product="5'-3' exonuclease"
           /protein_id="CAD02798.1"
           /db_xref="GI:16504052"
           /db_xref="SPTREMBL:Q82433"
           /translation="MLASPRRECTPFRFPRENTVAAHLIIVDANLIRIRIHAVQSP
           C/TCOALDQULIHSQTHAVAFDDDASSSGRGGHRTDYKAGRPEDDLENEF
           ALPAEFGRCVCMASDGNADLAATLAKTBAGQATIVSTDKYGLSEGLTI
           RDTFKRMADLAPTEKEFGVLPQLPDYMWLAGISSKPGVAGIGKXSTQLVQPO
           NLGILYAHLDVEPEKMKLETHKEMAFICRDIAIQTDLIDGNLQQLRVR"
misc_feature 3201..3947
           /gene="STRY3111"
           /note="pfam match to entry PF01367 5_3-exonuclease, 5'-3'
           exonuclease, score 120.40, E-value 3.3e-32"
           complement(4042..5190)
           /gene="STRY3112"
           /note="fucp"
           complement(4042..5190)
           /gene="STRY3112"
           /note="Fasta hit to EUTG_ECOLI (395 aa), 40% identity in
           372 aa overlap
           Fasta hit to Adh2_ECOLI (383 aa), 44% identity in 371 aa
           overlap
           Orthologue of E. coli fucp (FUCO_ECOLI); Fasta hit to
           FUCO_ECOLI (383 aa), 91% identity in 382 aa overlap"
           /codon_start=1
           /transl_table=11
           /product="1,2-propanediol oxidoreductase (lactaldehyde
           reductase)"
           /protein_id="CAD02799.1"
           /db_xref="GI:16504053"
           /db_xref="SPTREMBL:Q82432"
           /translation="MANRMILNETAWMGKGAVALDTDEVTREGYHALLIYTDKTLVOC
           CKVADKTSRMDAAGLAWIYAGVLPNPTLISVQGLFPTQSGADYLIATIGSGSPQT
           CKVAGIISNNPEADVRSLEGLSPTKRPSPVIMAIPTAGIAAEVITINIVTBEEKR
           KFCVCDPHDIPQAAFIADADWMDKMPALPKATGDADALTHAIEGVTAAVALDTHAI
           KALEIIGALRGAAVAGEKEGEMALQOQYVAKMGFSNVGGLVGMHAPGAFNTPH
           GVANAAILPHVMRFNAGSTNEKFRDIAIRANGVKEGSLSEARRAANAEEAFTLRDVG
           IPILRDVGVRKEDIIPALQAAFDVCTCGPREASLADIIVELVHTAW"

```

```

misc_feature complement(4069..5169)
           /gene="STRY3112"
           /note="pfam match to entry PF00465 Fe-ADH, Iron-containing
           alcohol dehydrogenases, score 655.70, E-value 2.4e-193"
           complement(4354..4416)
           /gene="STRY3112"
           /note="PS00060 Iron-containing alcohol dehydrogenases
           signature 2"
           complement(4585..4671)
           /gene="STRY3112"
           /note="PS00913 Iron-containing alcohol dehydrogenases
           signature 1"
           complement(5207..5854)
           /gene="STRY3113"
           /note="fucA"
           complement(5207..5854)
           /gene="STRY3113"
           /note="Fasta hit to YGBL_ECOLI (212 aa), 34% identity in
           179 aa overlap
           Orthologue of E. coli fucA (FUC_A_ECOLI); Fasta hit to
           FUC_A_ECOLI (215 aa), 92% identity in 215 aa overlap"
           /codon_start=1
           /transl_table=11
           /product="fuculose-1-phosphate aldolase"
           /protein_id="CAD02800.1"
           /db_xref="GI:16504054"
           /db_xref="SPTREMBL:Q82431"
           /translation="MERNRLARQIIDTCLMTRLGLNQGTAQNVSVYOGMLITPBTG
           IPEKLTESHIVFIDADQGEQGLPSSSEMRHMAVQTRPPANAIVHNAVHCAVAS
           IINRPIPAIHVMIAAGNSIPCAPATGTRELSHVAVLAKNRKATLLQHGGLIAC
           ENIDKALMLAHREVIQAQVLTALVDPVPLVDEALVIALEKRYGLRIEE"
misc_feature complement(5219..5845)
           /gene="STRY3113"
           /note="pfam match to entry PF00596 Aldolase II, Class II
           Aldolase and Adducin N-terminal domain., score 337.10,
           E-value 2e-97"
           6410..7726
           /gene="STRY3115"
           /note="fucp"
           6410..7726
           /gene="STRY3115"
           /note="Orthologue of E. coli fucp (FUCP_ECOLI); Fasta hit
           to FUCP_ECOLI (438 aa), 92% identity in 438 aa overlap"
           /codon_start=1
           /transl_table=11
           /product="L-fucose permease"
           /protein_id="CAD02801.1"
           /db_xref="GI:16504055"
           /db_xref="SPTREMBL:Q82430"
           /translation="MGNNTTIQTSFRAVDAEBSKSRYYIIPFALCSLFFLVANNL
           NDILLPOFOAFLTNFOAGLIQSAFVGFVPIPIAGILMKLSKAGITGLFLYA
           FGAALFWPAEINNNYTLFLIGLFIAGAGCCELTANPNPVALGSPGGHFRNLNLOT
           FNSFGAIIIVAFEGOSLILSNVPHOSQALDKMTDPDLSAKSLVSVQTPVMIIVAI
           VIVVALLIMLTKFPALOSDDSDPAQOSSELSLSLRIRIRHMMWAVLAQFCVGAQTA
           CMSYLIRIYALIEIPGMTPGFPAANYLTGTWCCEFTGFTGTLISRAPHFVLAAYAL
           AMULCISAFSGGHIILALTLCSAFMSIOYPIIFSLGKNLQDRTKYSSTFLVMTII
           GGGIVTPWVGFSVDAAGKIPTAELVPALCFAVIFIRRRSQATV"
gene      7758..9488
           /gene="STRY3116"
           /note="fucI"
           7758..9488

```

Alignment Scores:

Pred. No.:	Score:	Percent Similarity:	Best Local Similarity:	Query Match:	DB:
0.00535	233.50	40.19%	24.44%	13.92%	1
230050	76	99	87	9	

US-10-018-706-2 (1-322) x AL627277 (1-230050)
 Qy 18 LeuGlyLeuIlePheGlyValIleThrThrCysIleLeuAlaGlyCysAlaSerLysPro 37


```

Db      84819 CTGGGATTATGCGGTGCTGATATGACAGCGGCTGCTGCGGGGTGT----- 84772
Qy      38 ThrThraSerThrSerGlySerHisArgThrSerGlyGlyLeuAla 57
Db      84771 -----TCCAGTATATCCGGCTCAGGAAAGTAT-----TCCGGCTCC----- 84736
Qy      58 IleGlySerGlnValIleThrAspSerGlnGlyValProAsnArgTyrGlnValIleGln 77
Db      84735 -----GTCATACCGTTTAAGCGC 84718
Qy      78 GlyAspThrValSerIysIleAlaGlnArgTyrGlyLeuAsnTyrArgGluIleGlyHis 97
Db      84717 GGGGATACGTTGATGATATATCGCGGTACGGGAACCGCTGAAGAAGATCGCCCGA 84658
Qy      98 IleAsnAsnLeuAsnSerSerTyrThrIleTyrThrGlyGlnThrPheLeuThrIleTyrPse 117
Db      84657 CTGAATGGCATCTTCTCGCTTATCTATTAGAGTGGGAGCGAAGTAAAGTG--CGC 84601
Qy      118 GlyAspLeuLys-----ValArgGluArgSerIleSerSerGlyValAsnThr 133
Db      84600 GCGACGGCCAAAGAGTAGCAGTTCAACACGCAAAACGCAATAAACAGCAACAAACT 84541
Qy      134 AlaHisThr---ProSerProValAlaValGlnSerSerArgProProValGlnGlnHis 152
Db      84540 GCTGCGGTACAGACCGTATCTCGGTGCGCAAAATCTCTGCGCGCGCGGTGGGCGACGCT 84481
Qy      153 ProAlaValGlnLysProThrProProValValValIleValLysLysProThrPro 172
Db      84480 TGTGGGTATGGCGCTTAACGGTAAAGTATTTTA----- 84445
Qy      173 ProValValGlnGlnProAlaProValAlaProProValThrGlnAlaProPheAlaThr 192
Db      84444 -----CTTATTCACCC 84433
Qy      193 GlySerSerGlyValMetGlnPheArgTyrProValGlyValaThrAsnProValArg 212
Db      84432 GCGGAGGCGCGA----- 84421
Qy      213 ArgPheGlyThrAlaThrValAlaGlySerThrValThrSerAsnGlyMetTyrPheSer 232
Db      84420 -----AATAAGGGATCGATATTGCGC 84400
Qy      233 GlyArgAspGlyAspLeuIleAsnAlaSerAsnAlaGlyThrValIleGlnAlaAspHis 252
Db      84399 GCGCGCGCGGAGACCGCTGTTATGCGGCGGAGCGGTAAGTTGCTATGTCGGGAC 84340
Qy      253 AsnMetAspGlyAlaSer-----IleValIleGlnHisThrAsnGlyPheValSerSer 270
Db      84339 CAGCTCGCGGTTACCGCATTCATCATGATTAAAGCAATGAAGTTACATTACGCGCT 84280
Qy      271 TyrIleHisIleIleAspAlaGlnValLysThrGlyAspThrValArgThrGlyGlnArg 290
Db      84279 TACGCTCATATGACACCATGCTGTTATATATGACCAAGCCGTTAAAGCGGCGCAAAA 84220
Qy      291 IleAlaSerMetLysAsnGlnProSerGlyValAlaLeuPheGlnPheArgIleSerArg 310
Db      84219 ATCGCTACGATGGGACGACGAGATGCCGATCCGTAAGACTTACATTCCGATTCGCTAC 84160
Qy      311 AsnGlyValTyrValAspProLeuThrValIleu 321
Db      84159 CCGGCGACGCGCATCATCGCTGCGCTTATCTG 84127

```

RESULT 19
AF260132
LOCUS AF260132 1522 bp DNA linear BCT 16-MAY-2000
DEFINITION Pseudomonas putida putative protein-L-isoaspartate
O-methyltransferase (pcm) gene, partial cds; and NlpD (nlpD) gene,
complete cds.
VERSION AF260132.1 GI:7839529
KEYWORDS
SOURCE Pseudomonas putida.

ORGANISM Pseudomonas putida
Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
Pseudomonas.

REFERENCE 1 (bases 1 to 1522)
Ojangu, E.-L., Tover, A. and Kivisaar, M.
Sequence of Pseudomonas putida nlpD gene (complete sequence) and
ORF similar to E. coli pcm gene (partial sequence)

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1522)
Ojangu, E.-L., Tover, A. and Kivisaar, M.
Direct Submission
Submitted (24-APR-2000) Genetics, Institute of Molecular and Cell
Biology at Tartu University, Riia 23, Tartu 51010, Estonia
Location/Qualifiers
1..1522

FEATURES
source
/organism="Pseudomonas putida"
/strain="KT2440"
/db_xref="taxon:303"
<1..326
/gene="pcm"
<1..326
/gene="pcm"
/note="pcm"
/note="methyltransferase"
/codon_start=3
/transl_table=11
/product="putative protein-L-isoaspartate
O-methyltransferase"
/protein_id="AF70310.1"
/db_xref="GI:7839530"
/translation="RRSIDLHDAKERLVELNRYVFMGDCGEMPALAPYNGITV
TAVAPVPOALLDQALPAGGRMVIPVPAEAOQMLIVREHGFSSRVILGAVFVPL
NGPLA"
631..1440
/gene="nlpD"
631..1440
/gene="nlpD"
/note="nlpD"
/note="lipoprotein"
/codon_start=1
/transl_table=11
/product="NlpD"
/protein_id="AF70311.1"
/db_xref="GI:7839531"
/translation="MGSLVGCSSTSSNSARVVDNRNNTYKRPATVSGOYIVKPGDTL
FSIAFVGMVDYKELARNGIOAPYRTIRPGAIRFSSGSSRTTVVSPSSSTTVR
RPVGRATRAPASTSKPATSPAPAVVATVPAPBAVAGVMTVPANGVILGRASNS
LNKGIDIALGQPVTPASDGAIVAGSGRGELITIHSDTYSAVGHNRLLVR
EGQVRAQGSILKMYVGTDRVKLHFIRROETRRSTPVPATSLTVPSLPL"

BASE COUNT 274 a 438 c 486 g 324 t

ORIGIN

Alignment Scores:
Pred. No.: 2,42e-05 Length: 1522
Score: 231.50 Matches: 82
Percent Similarity: 41.96% Conservative: 38
Best Local Similarity: 28.67% Mismatches: 101
Query Match: 13.86% Indels: 65
DB: 1 Gaps: 12

US-10-018-706-2 (1-322) x AF260132 (1-1522)

Qy 20 leuIlePheGlyValIleThrThrCysIleLeuAlaGlyCysAlaSerLysProThrTyr 39
Db 613 CTGGTATTCATGATGGCCATGGCTGCTTCTGCTGGGTTGCTCAAGC----- 660

Qy 40 AsnSerThrSerGlySerGlySerHisArgThrSerGlySerGlyGlyLeuAlaIleGly 59
Db 661 -----ACCAGCAGCAAC-----AGC 675

Qy 60 SerGlnValIleThrAspSerGlnGlyValProAsnArg----- 72
Db 676 GCGCGGTGCTGTCACCGCAACACACCGGCCGCGCGGTGATCTTGGGCAA 735

Qy 73 TyrGlnValIleGlnGlyAspThrValSerLysIleAlaGlnArgTyrGlyLeuAsnTyr 92

[illegible]

REFERENCE 3 (bases 1 to 2609)
AUTHORS Seshadri, R. and Samuel, J. E.
TITLE Direct Submission
JOURNAL Submitted (01-JUN-2000) Medical Microbiology and Immunology, Texas A and M Health Science Center, 407 Reynolds Medical Bldg., College Station, TX 77843-1114, USA
REMARK Sequence update by submitter
COMMENT On Jun 1, 2000 this sequence version replaced gi:7340284.
FEATURES
source
1..2609
/organism="Coxiella burnetii"
/db_xref="taxon:777"
/note="nine mile phase I"
1..777
/note="function unknown"
/codon_start=1
/transl_table=1
/product="Surs"
/protein_id="AAF73516.1"
/db_xref="GI:8141682"
/translation="MKRTATPKURLISNDGQVYAKGLATLADLGEVDVYAPRR
NRGSASNTLTNAPLHAKNLNENGI SVETPTDCHLAIATGVLPENPDVYAGINAGP
NLGDVWVSGTVAAMGRLGLPALVSGELGPIRYETPAKVYOLIQLEKDPLE
PSTLTINVPDLPYEELKGEFVRLGRHAEPTIRQIDRGHPTIYVGAAGPEQDSG
GTGDFPAMNHHCVSITPLRVDLTHYEAFLDQLAAGVKLEEM"
779..1471
/codon_start=1
/transl_table=1
/product="Lipoprotein Precursor NlpD"
/protein_id="AAF73517.1"
/db_xref="GI:8141683"
/translation="MKRCIVIIIAFLICGCVESKNFAPVNGWLQPKARLSYKVKQ
GDTYSIAFMAGFLGRLAALANRSPRIEACGLQMTIILPRANHSGRAPAPRMQ
SPPLQVAVRMWPAKRLIGYYSVGAAGNGIYIASHGVAADGAVVYAGAGVQ
RGYNLTIVGHNTIYLSAVAFNKKVLKESGRVAGQKIAMGRTNSGRVWLHFEIRH
NGQVNFRLYIS"
1451..2609
/note="global regulator of gene expression"
/codon_start=1
/transl_table=1
/product="stress response sigma factor Rpos"
/protein_id="AAF81172.1"
/db_xref="GI:7402865"
/translation="MKRTTKTKTKRAKTKIKRSKRIKTKAKSRPKPKIKASDHG
LIFKATKETTERKEDAEALANKATKRAKTRRSDDPTQIYLRGROPILNAKEIKTK
ARRVHKDPPKARKOMTEANLNLVYKIRAHYKRLPLDLIEENLGLTRPDEPDE
RGFEFSTVATWIRIQTIERAIMNSKTVLPFHVKEINLYLRKAKIQGVDEAHF
EDVALLIDKPEVQETIRIMDLSPASITDPLISEDQSLVDTLADNNIDPARLQNN
DLQHIRRWLAQLDERREVVILRFGLHNEKGTLEAVGAVGLRERKROIQIDALC
QLRHILMEGTGSEVD"

ORIGIN 730 a 568 c 676 g 635 t

ALIGNMENT SCORES:
Pred. No.: 7.56e-05 Length: 2609
Score: 227.50 Matches: 82
Best Local Similarity: 38.61% Conservative: 35
Best Local Similarity: 27.06% Mismatches: 101
Query Match: 13.62% Indels: 86
DB: 1 Gaps: 9

US-10-018-706-2 (1-322) x AF244357 (1-2609)

QY 24 Valletthrrhrcysileleualgilycyvalaserlys-Prothrrtyraanserthse 43
Db 800 ATCATATTCGCTTTCCTTGTTCGCGCTTGTGTAAACCAAACTTTGGCGCTGTGTC 859
QY 43 rglyseratgyerhsaagthrserserglysercglycyleualaleeglyserglnvali 63
Db 860 AACGCGCTGACTTCAACGCAAGCAGCGGTTA-GGTTCT----- 895
QY 63 ethraspserglnglyvalprobsnarytrglnvallysgelnglyasgptnvalserly 83

Db	896	-----TATGGGCTCAAGACGGGGATACATTACT	927
Qy	83	sllelaaglaargtyrglyleuaasrtparglulilegylhslleasnaasleuase	103
Db	928	GATTGCCCGGGCGCTGGCTGATATTCAGACCTTACCGCGCCCAATCGTTTAAGCCC	987
Qy	103	rserlythrillelytrhnglyglntprpleuthreutrpserglyaspleuylvalar	123
Db	968	TCCTTATGCAATTGACGGGGGCTCAACTTTACAAATG-----	1024
Qy	123	ggluargerlleserSerglyValaasnthrLahlethrProserProvalAlaValGl	143
Db	1025	-----ACGAAATTCCTCGAGGC-----GCGCAGCTTTCTGGCGCTTTTGGC-----	1066
Qy	143	nserSerarproProvalGlnGlnHleProalavalGlnlySProthrProProvalva	163
Db	1066	-----	1066
Qy	163	lvalValylslySProthrProthrProProvalValGlnGlnHleProalAlaPr	183
Db	1067	-----GCGGCTCTTACATGATGCAATCTCCCGCGCGCTCGACGC	1104
Qy	183	oProvalThrGluAlaProPheAlaThrGlySerSerSerglyValMetGlnPheArgTyPr	203
Db	1105	C-----GTCGCCCATTCGCGCATGCGCC	1125
Qy	203	ovalGlyAlaThrAsnProvalValArgArgPheGlyThrAlaThrVal-----AlaGl	221
Db	1126	TGCCAAAGAGG-----CGCCCTCATTCGGATTAATTCGTCTGCGCATGCGCGG	1170
Qy	221	ySerThrAlaThrSerAsnGlyMetTrpPheSerGlyArgArgPglYaspleuileasna	241
Db	1171	G-----AATCAAGAGATCAATATTTCAGAGCATTAATGATGAGGCGGCTCCGCGC	1218
Qy	241	aserasnaaglyThrValileGlnAlaAspHisasMetaspglyValaser-----Il	259
Db	1219	CGCTTCCCATGAGACTTGTATATAGTCGTCGCTGTACAGAGTTATGTAATCTTAT	1278
Qy	259	evalileGlnIsthraasnglyPheValSerSerTyrlleHsllelysaaspaalaglnva	279
Db	1279	AATGTAAAGCATACGAATACTTATTAAAGTCCTATCGTTCAACAAACGGGTCCTCGT	1338
Qy	279	llyethrGlyAspThrValArgThrGlyGlnArgyleaIsasermetylsaenglnProse	299
Db	1339	CAAGGAAAGTGTAGGCTTGTGTGTGACAGAAAGTTCCAGAGATGGGCGGTACTACTC	1398
Qy	299	rglyAlaAlaLeuPheGlnPheArgIleSerArgAsnGlyValTyTrpAspProleuth	319
Db	1399	AGGACGAGTAATCTTCAATTGAAATCCGCCGAATAGCCAGCAGTGAATCCGCTTCG	1458
Qy	319	rValleu	321
Db	1459	GTAAGTTG	1465
RESULT 21			
AE005516/c			
LOCUS	AE005516	9360 bp	DNA linear BCT 21-MAR-2001
DEFINITION	Escherichia coli O157:H7 EDL933 genome, contig 3 of 3, section 135		
ACCESSION	AE005516		
VERSION	AE005516.1		
KEYWORDS	gi:12517379		
ORGANISM	Escherichia coli O157:H7 EDL933.		
SOURCE	Escherichia coli O157:H7 EDL933		
REFERENCE	1 (bases 1 to 9360)		
AUTHORS	Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D., Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A., Poefai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L., Grobbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamoudis,K., Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C., Welch,R.A. and Blattner,F.R.		

TITLE	Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
JOURNAL	Nature 409 (6819), 529-533 (2001)
MEDLINE	21074935
PUBMED	11206551
REFERENCE	2 (bases 1 to 9360)
AUTHORS	Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D., Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A., Poefai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L., Grobbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamoudis,K., Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C., Welch,R.A. and Blattner,F.R.
TITLE	Direct Submission
JOURNAL	Submitted (22-OCT-2000) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
FEATURES	Location/Qualifiers
SOURCE	1. 9360
	/organism="Escherichia coli O157:H7 EDL933"
	/strain="EDL933"
	/serotype="O157:H7"
	/db_xref="taxon:155864"
	/note="enterohemorrhagic"
misc_feature	<1..923
	/note="O-island #115; Region of the EDL933 chromosome not homologous to E. coli K-12 MG1655"
gene	complement (204..383)
	/gene="Z4199"
CDS	complement (204..383)
	/gene="Z4199"
	/function="orf; Unknown function"
	/note="No significant matches"
	/codon_start=1
	/transl_table=11
	/protein_id="AA657991.1"
	/db_xref="GI:12517380"
	/translation="MINKLLAVLIGLVVSTLIIFSEKVTYRLFAAITGTWPLSLPSISLMIRKSD"
gene	complement (450..1478)
	/gene="Z4200"
CDS	complement (450..1478)
	/gene="Z4200"
	/function="orf; Unknown function"
	/note="Residues 87 to 185 of 342 are 97.97 pct identical to residues 1 to 99 of 99 from Escherichia coli K-12 Strain MG1655: B2862"
	/codon_start=1
	/transl_table=11
	/protein_id="AA657992.1"
	/db_xref="GI:12517381"
	/translation="MMADVDEBEKFPQIMNKHNDALGLYDILLIQAINGTINAEWMQKGRLLKVALILAVLDKAFNYWEDSGAMEDEARLTSSVALVTSGIERLSNLISKDSEFVSDILREAKVNEIDETLSTIRLHLIDKGYERITLQDLGSESPYDEKHYEADAAALNVIYPAVNLKINTREKQVLIYKLAGPYGIRKYEKONTOSAFNWDITDTONSHAREKRSFIPSTAEWEPFDSVYAKSAIYKSEKREKYLNDVQFNMRISQITGEMNIGANGRSVPEMALPESNYIHKSGTTHAEPSPITPLNWSRASMTLMEKSNLIMDEGIR"
	complement (1492..2097)
gene	/gene="Z4201"
CDS	complement (1492..2097)
	/gene="Z4201"
	/function="orf; Unknown function"
	/note="Residues 1 to 201 of 201 are 99.00 pct identical to residues 1 to 201 of 278 from Escherichia coli K-12 Strain MG1655: B2863"
	/codon_start=1
	/transl_table=11
	/protein_id="AA657993.1"
	/db_xref="GI:12517382"
	/translation="WVIYAFNKKLMEYFMKGSALTLLAGIPSCGTQATGAEVTSSEVNIINLTAAADKSYLSLNPDKYNYRLIHSAKIQEIKSHYTKDEIQGLITTE NTRKLTTEKPEKSTYAFEDDKYAAAEHYDAVWLRDSLMGYALVSDQNSVAAK KVLTLMDWVSTDQIKRMQDIDISNRKRDGIFGQNNAYHI"
	2198..2275
	/note="O-island #116; Region of the EDL933 chromosome not

misc_feature

2198. .2275
/note="O-island #116; Region of the EDL933 chromosome not

gene homologous to E. coli K-12 MG1655
complement (2289. .2362)
/gene="glyt"
/note="24202"
complement (2289. .2362)
/gene="glyt"
/product="tRNA-Gly"
/note="anticodon: CCC"
/function="tRNA: tRNA"
complement (2487. .3158)
/gene="24203"
complement (2487. .3158)
/gene="24203"
complement (2487. .3158)
/function="putative structure; Not classified"
/note="Residues 1 to 192 of 223 are 98.5 pct identical to
residues 22 to 213 of 259 from Escherichia coli K-12
Strain MG1655"
/codon_start=1
/transl_table=1
/product="putative 11poreprotein"
/protein_id="AAG57994.1"
/db_xref="GI:12517383"
/transl_table="MMLLSVGLLAGCGSKSPDTGYSGSVTVKRGDTLYRISRT
GTSVEKLRLNIGISPTPIEYGOGLKLGAGSSSTSRKSTAKSTTKTASVTPSSAPV
GSPVPGVORLNIISPTPIEYGOGLKLGAGSSSTSRKSTAKSTTKTASVTPSSAPV
RSGYGLNIMKISSEDIYTAVANDTMLVNNQSGXSGEMAKNRHGHGCGICSPAPFPL
PQNGN"

gene 3652. 5910
/gene="24205"
3652. 5910
/gene="24205"
/function="orf; Unknown function"
/note="Residues 1 to 752 of 752 are 99.46 pct identical to
residues 1 to 752 of 752 from Escherichia coli K-12 Strain
MG1655"
/codon_start=1
/transl_table=1
/product="orf, hypothetical protein"
/protein_id="AAG57995.1"
/db_xref="GI:12517384"
/transl_table="MRVDAIAKVTGRRARYTDDYVMAQMCYAKVRSPIAHGVAISND
EARSLLPGVALFTMEVDVPIEFPAVGAHWLIDENKRDADRALTBHVAHDAVAI
VVARBELTKEKAOLVSIEMBELPVITTPBEAALLDEAPAHINGNLLKOSTNSTNQV
QTLIDADVQVGHQYQTPVIOHOMESVTSLLAMDDSDSITVSTQIPHIVRVVSGA
LDIPMSCAVITKTFVGGFGFNKQDVLSEPRMAAFSTSKGIPVVSLSRECFCLART
RLAFTIDCGMVGNDGTGLKYSVDLVSNTGAIVSHGISAGCNKTVAYLPCACAY
SRTKCTNIPSAAGARGYGAPQVPAVFSMLDDAATLALGIDPVEIRLRNARSQDAP
LNGKATYSAGPECLKEGRKIFEMEKRAECQNOQGNLRGVGAFCFSYTSNTMPV
FADGALNNQDGTINVOGATIEIGQADTFVFSQVMAETVGPVSDVRSITDDTDT
PPDQGFARSGOYVAAPALRSALLLKKEKIIAHAAMVHGSAMNLTLLKGIYVLEP
BEPLMSKLTAMDAPYHBERGQGLSASEISITKTNPPAFGTFVDLTVDALCKVTN
RLNWDGSHLINDPLLAGGVGAGMGWQMGWIGALPEEMITDKSGSVNPNMLLYKPT
MDLLPLESAFLEINERPOSAVGHKSLGSPPIIPAAALIRNVAKGTVAINTLPLPDK
RLRPEPHLAGLI"

gene 5921. 6799
/gene="ygeet"
/note="22366"
5921. 6799
/gene="ygeet"
/function="putative enzyme; Not classified"
/note="Residues 1 to 292 of 292 are 96.97 pct identical to
residues 1 to 292 of 292 from Escherichia coli K-12 Strain
MG1655"
/codon_start=1
/transl_table=1
/product="putative dehydrogenase"
/protein_id="AAG57996.1"
/db_xref="GI:12517385"
/transl_table="MFDPAASYRAATTLADAINLADNPQALLAGTDDVLIOLHHND
RRHYIADINHLKIGTLTAEQSLRIGSATTFQXLEDSTQRLPALCAASSLAG
POIRNATVATCNGATGADGATPTLLYDAKLHISPRGRFVPINGFTGQKSL
EHDELIVAFPEPQEHVGSAPFYARMADNDISTGCAACGLDNGENSERLAFG
VAFPTIRCOHAEQIAONAPNLQITLAEIISSEVADVPAPRSMAKSEPRHILQWT

gene 5921. 6799
/gene="ygeet"
/note="22366"
5921. 6799
/gene="ygeet"
/function="putative enzyme; Not classified"
/note="Residues 1 to 292 of 292 are 96.97 pct identical to
residues 1 to 292 of 292 from Escherichia coli K-12 Strain
MG1655"
/codon_start=1
/transl_table=1
/product="putative dehydrogenase"
/protein_id="AAG57996.1"
/db_xref="GI:12517385"
/transl_table="MFDPAASYRAATTLADAINLADNPQALLAGTDDVLIOLHHND
RRHYIADINHLKIGTLTAEQSLRIGSATTFQXLEDSTQRLPALCAASSLAG
POIRNATVATCNGATGADGATPTLLYDAKLHISPRGRFVPINGFTGQKSL
EHDELIVAFPEPQEHVGSAPFYARMADNDISTGCAACGLDNGENSERLAFG
VAFPTIRCOHAEQIAONAPNLQITLAEIISSEVADVPAPRSMAKSEPRHILQWT

gene 5921. 6799
/gene="ygeet"
/note="22366"
5921. 6799
/gene="ygeet"
/function="putative enzyme; Not classified"
/note="Residues 1 to 292 of 292 are 96.97 pct identical to
residues 1 to 292 of 292 from Escherichia coli K-12 Strain
MG1655"
/codon_start=1
/transl_table=1
/product="putative dehydrogenase"
/protein_id="AAG57996.1"
/db_xref="GI:12517385"
/transl_table="MFDPAASYRAATTLADAINLADNPQALLAGTDDVLIOLHHND
RRHYIADINHLKIGTLTAEQSLRIGSATTFQXLEDSTQRLPALCAASSLAG
POIRNATVATCNGATGADGATPTLLYDAKLHISPRGRFVPINGFTGQKSL
EHDELIVAFPEPQEHVGSAPFYARMADNDISTGCAACGLDNGENSERLAFG
VAFPTIRCOHAEQIAONAPNLQITLAEIISSEVADVPAPRSMAKSEPRHILQWT

gene 5921. 6799
/gene="ygeet"
/note="22366"
5921. 6799
/gene="ygeet"
/function="putative enzyme; Not classified"
/note="Residues 1 to 292 of 292 are 96.97 pct identical to
residues 1 to 292 of 292 from Escherichia coli K-12 Strain
MG1655"
/codon_start=1
/transl_table=1
/product="putative dehydrogenase"
/protein_id="AAG57996.1"
/db_xref="GI:12517385"
/transl_table="MFDPAASYRAATTLADAINLADNPQALLAGTDDVLIOLHHND
RRHYIADINHLKIGTLTAEQSLRIGSATTFQXLEDSTQRLPALCAASSLAG
POIRNATVATCNGATGADGATPTLLYDAKLHISPRGRFVPINGFTGQKSL
EHDELIVAFPEPQEHVGSAPFYARMADNDISTGCAACGLDNGENSERLAFG
VAFPTIRCOHAEQIAONAPNLQITLAEIISSEVADVPAPRSMAKSEPRHILQWT

gene 5921. 6799
/gene="ygeet"
/note="22366"
5921. 6799
/gene="ygeet"
/function="putative enzyme; Not classified"
/note="Residues 1 to 292 of 292 are 96.97 pct identical to
residues 1 to 292 of 292 from Escherichia coli K-12 Strain
MG1655"
/codon_start=1
/transl_table=1
/product="putative dehydrogenase"
/protein_id="AAG57996.1"
/db_xref="GI:12517385"
/transl_table="MFDPAASYRAATTLADAINLADNPQALLAGTDDVLIOLHHND
RRHYIADINHLKIGTLTAEQSLRIGSATTFQXLEDSTQRLPALCAASSLAG
POIRNATVATCNGATGADGATPTLLYDAKLHISPRGRFVPINGFTGQKSL
EHDELIVAFPEPQEHVGSAPFYARMADNDISTGCAACGLDNGENSERLAFG
VAFPTIRCOHAEQIAONAPNLQITLAEIISSEVADVPAPRSMAKSEPRHILQWT

gene 5921. 6799
/gene="ygeet"
/note="22366"
5921. 6799
/gene="ygeet"
/function="putative enzyme; Not classified"
/note="Residues 1 to 292 of 292 are 96.97 pct identical to
residues 1 to 292 of 292 from Escherichia coli K-12 Strain
MG1655"
/codon_start=1
/transl_table=1
/product="putative dehydrogenase"
/protein_id="AAG57996.1"
/db_xref="GI:12517385"
/transl_table="MFDPAASYRAATTLADAINLADNPQALLAGTDDVLIOLHHND
RRHYIADINHLKIGTLTAEQSLRIGSATTFQXLEDSTQRLPALCAASSLAG
POIRNATVATCNGATGADGATPTLLYDAKLHISPRGRFVPINGFTGQKSL
EHDELIVAFPEPQEHVGSAPFYARMADNDISTGCAACGLDNGENSERLAFG
VAFPTIRCOHAEQIAONAPNLQITLAEIISSEVADVPAPRSMAKSEPRHILQWT

gene 5921. 6799
/gene="ygeet"
/note="22366"
5921. 6799
/gene="ygeet"
/function="putative enzyme; Not classified"
/note="Residues 1 to 292 of 292 are 96.97 pct identical to
residues 1 to 292 of 292 from Escherichia coli K-12 Strain
MG1655"
/codon_start=1
/transl_table=1
/product="putative dehydrogenase"
/protein_id="AAG57996.1"
/db_xref="GI:12517385"
/transl_table="MFDPAASYRAATTLADAINLADNPQALLAGTDDVLIOLHHND
RRHYIADINHLKIGTLTAEQSLRIGSATTFQXLEDSTQRLPALCAASSLAG
POIRNATVATCNGATGADGATPTLLYDAKLHISPRGRFVPINGFTGQKSL
EHDELIVAFPEPQEHVGSAPFYARMADNDISTGCAACGLDNGENSERLAFG
VAFPTIRCOHAEQIAONAPNLQITLAEIISSEVADVPAPRSMAKSEPRHILQWT

gene 5921. 6799
/gene="ygeet"
/note="22366"
5921. 6799
/gene="ygeet"
/function="putative enzyme; Not classified"
/note="Residues 1 to 292 of 292 are 96.97 pct identical to
residues 1 to 292 of 292 from Escherichia coli K-12 Strain
MG1655"
/codon_start=1
/transl_table=1
/product="putative dehydrogenase"
/protein_id="AAG57996.1"
/db_xref="GI:12517385"
/transl_table="MFDPAASYRAATTLADAINLADNPQALLAGTDDVLIOLHHND
RRHYIADINHLKIGTLTAEQSLRIGSATTFQXLEDSTQRLPALCAASSLAG
POIRNATVATCNGATGADGATPTLLYDAKLHISPRGRFVPINGFTGQKSL
EHDELIVAFPEPQEHVGSAPFYARMADNDISTGCAACGLDNGENSERLAFG
VAFPTIRCOHAEQIAONAPNLQITLAEIISSEVADVPAPRSMAKSEPRHILQWT

gene 5921. 6799
/gene="ygeet"
/note="22366"
5921. 6799
/gene="ygeet"
/function="putative enzyme; Not classified"
/note="Residues 1 to 292 of 292 are 96.97 pct identical to
residues 1 to 292 of 292 from Escherichia coli K-12 Strain
MG1655"
/codon_start=1
/transl_table=1
/product="putative dehydrogenase"
/protein_id="AAG57996.1"
/db_xref="GI:12517

gene
KVI9.SEVAAGAAGLQ
6796. .7275
/gene="Z4207"
6796. .7275
/gene="Z4207"
/function="putative enzyme; Not classified"
/note="Residues 1 to 159 of 159 are 99.37 pct identical to
residues 1 to 159 of 159 from Escherichia coli K-12 Strain
MG1655. B2868"
/codon_start=1
/transl_table=1
/product="putative dehydrogenase"
/protein_id="AAG57997.1"
/db_xref="GI:12517386"
/translation="MNSRTTTECTINGMPQLHAVPTPLSELRLREGLISYKQCC
CVBGCACVLVDGTALDCLYLAAAEKEITRLSEARAGKLSHVQAAVAGAVQ
CGFPGELIMATYATMAAKPERKPLTTEIRRGAGNLCRCGTGYMTVNTLDEKTK"
complement (7315. .9093)
/gene="ygev"
/note="Z4208"
complement (7315. .9093)
/gene="ygev"
/function="putative regulator; Not classified"
/note="Residues 1 to 592 of 592 are 99.32 pct identical to
residues 1 to 592 of 592 from Escherichia coli K-12 Strain
MG1655. B2869"
/codon_start=1
/transl_table=1
/product="putative transcriptional regulator"
/protein_id="AAG57998.1"
/db_xref="GI:12517387"
/translation="MELATQSVIMOTIOFPARMYASVLOEVEIDENYEVVAG
TGAERGLRQISGNSRLCHVLETRKXVDSRPPDLCEGDSKENCERKFLGPR
VILIDRCVGVISLIAYTHPOEHSIDRLSEKPVDSVIRITFVSKLELDGPGNLSK
IFAMTINMGCVIYVDSRQVDFPVNTAKTGVYNNIIGDIPRLPPLPSNFTPH
GHMHITVSDKSLIIGCLHNTQGRLEPLMAHPOSITSSVANAPDEPIELQYGC
RVMQLKRLISRIAPSSVAVVSGTGKVVARAIIKLSGRNRFIINCAIIE

Alignment Scores:
Pred. No.: 0.000435 Length: 9360
Score: 225.00 Matches: 78
Percent Similarity: 39.49% Conservative: 46
Best Local Similarity: 24.84% Mismatches: 103
Query Match: 13.47% Indels: 87
DB: 1 Gaps: 8

US-10-018-706-2 (1-322) x AEO05516 (1-9360)

QY 16 LysArgLeuGlyLeuAlaIlePheGlyValAlaIleThrThrCysIleLeuAlaGlyCysAlaSer 35
Db 3173 AAATCTGGGGTATACGTGATGATGTTTATTCGCTTGACTGCTGTTTGGGGGGCTG----- 3120
QY 36 LysProThrTyraSerThrSerGlySerGlySerHisArgThrSerGlySerGlyGly 55
Db 3119 -----TCGGGTAGCAAAATCATCCGATCAGAGAACGTATTCGGC 3081
QY 56 LeuAlaIleGlySerGlnValIleIleThrAspSerGlnGlyValProAsnArgTyGlyVal 75
Db 3080 -----TCCGTTTACACCGTG 3066
QY 76 LysGlnGlyAspThrValSerIleValIleAlaGlnArgTyGlyLeuAsnTrpArgGlnIle 95
Db 3065 AAACGGGGGGATCGCTATTCGTAATTTCCGGACCCACGGGACACGCGTAAAGAAAGCTG 3006
QY 96 GlyHisIleAsnAsnLeuAsnSerSerTyThrIleTyThrGlyGlnTrpLeu 115
Db 3005 GCGCGACGAAAGCGCATTTCCCGCCCTTACACCATTAAGTTGCTCAAGAACTAAACTG 2946
QY 116 TrpSerGlyAspLeuTyS-----ValArgGluArgSerIleSerSerGlyVal 131
Db 2945 -----GGTGGGGCGAAAGAGTACGACGTACACGTAAATATACCGCCMAATCAACGACC 2892
QY 132 AsnThrAlaHis-----ThrProSerProValAlaValGlnSerSerArgProProValGln 150

Db	2891	AAAAACCGATCGTTACACCGCATCACCGGCTACCGGAATCTCTTGGCCGCGCAGTACGG	2832
Qy	151	GlnHisProAlaValGlnLysProThrProProValValValLysLysProThrPro	170
Db	2831	CAAGCTTGTTGGTTATGGCCAAACGACAGCAAGAAAGTTATCATG	2790
Qy	171	ThrProProValValGlnGlnProAlaProValAlaProProValThrGlnAlaProPhe	190
Db	2789	-----CCGTAAT	2784
Qy	191	AlaThrThrLysSerSerGlyValMetGlnPheArgTyrProValGlyAlaThrAsnProVal	210
Db	2783	TCGACAGACAGATGCGCGC	2766
Qy	211	ValAlaArgAlaPheGlyThrAlaThrValAlaIaGlySerThrValThrSerAsnGlyMetTrp	230
Db	2765	-----AATAAAGCATTTGAT	2751
Qy	231	PheSerGlyAlaArgAspGlyAspLeuIleAsnAlaSerAsnAlaGlyThrValIleGlnAla	250
Db	2750	ATTCACAGCTCCACGGGGGTACACCTATTTACCCCGGGGTGCAGGAAAGTGTATGTG	2630
Qy	251	AspHisAsnMetAlaPheGlyAlaSer-----IleValIleGlnHisThrAsnGlyPheVal	268
Db	2690	GGCAACACGCTCGCGGCTACGGTATCTCATCATGATTAACAACAGGAAGATTACATT	2631
Qy	269	SerSerTyrIleHisIleLysAspAlaGlnAlaLysThrGlyAspThr-ValAlaThrGln	288
Db	2630	ACGGCTTACGGCCATTAATGACACCATGCTGTGTAATATGGCCAAAGCKGTGAAGCTGG	2571
Qy	288	YGLNARGILEALASERMETLYASNGINPROSERGLYALALALEUPHEGLINPHEARGI	308
Db	2570	GCAAAATAATCCGACCATGCGGAGACACGAGATCGGCTGTGCGCTGCATTTCCAGAT	2511
Qy	308	eSerArgAsnGlyValTyrValaLeuProLeuThrValLeu	321
Db	2510	TCGTTACCGTGCAACGCAATTGATCGCTACGTTACTTG	2471
RESULT 22			
AL646063	207050 bp	DNA	linear
LOCUS			BCT 07-DEC-2001
DEFINITION	Ralstonia solanacearum GM1000 chromosome, complete sequence;		
ACCESSION	AL646063		
VERSION	AL646063.1		
KEYWORDS	GI:17428161		
SOURCE	Ralstonia solanacearum.		
ORGANISM	Ralstonia solanacearum		
REFERENCE	Bacteria; Proteobacteria; beta subdivision; Ralstonia group;		
AUTHORS	Ralstonia.		
	1 (bases 1 to 207050)		
	Salanoubat, M., Gentin, S., Attiguenave, F., Gouzy, J., Mengnot, S.,		
	Ariat, M., Billault, A., Broctier, P., Camus, D. C., Cattolico, L.,		
	Chandler, C., Choise, N., Claudel-Renard, C., Cunne, S., Demange, N.,		
	Gaspin, C., Lavie, M., Moisan, A., Robert, C., Saulin, W., Schlex, T.,		
	Signier, P., Thebaud, P., Whalen, M., Wincker, P., Levy, M.,		
	Weissenbach, J. and Boucher, C. A.		
TITLE	Genome sequence of the plant pathogen Ralstonia solanacearum		
JOURNAL	unpublished		
REFERENCE	2 (bases 1 to 207050)		
AUTHORS	Boucher, C. A.		
TITLE	Direct Submission		
JOURNAL	Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston		
	Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie		
	Moléculaire des Interactions Plantes-Microorganismes INRA-CNRS,		
	Bp27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean		
	Dausset-CBPB, 27 rue Juliette Dodu, 75010 Paris, France, IMGM CNRS		
	118 Route de Narbonne, F 31062 Toulouse Cedex 4, Genoscope and INRA		
	UR97, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France,		
	Laboratoire de Biométrie et Intelligence Artificielle INRA, Bp27,		
	F31326 Castanet-Tolosan Cedex. Laboratoire de Genétique Cellulaire		
	INRA, Bp27, F31326 Castanet-Tolosan Cedex		

COMMENT Christian.Boucher@toulouse.inra.fr
http://sequence.coulouze.inra.fr/R.solanacearum.html.

FEATURES
source location/Qualifiers

1. .207050
/organism="Ralsstonia solanacearum"
/strain="GMI1000"
/db_xref="taxon:305"
complement(332. .1561)
/gene="RSc1148"
/note="RS04604"
complement(332. .1561)
/gene="RSc1148"
/function="miscellaneous; hypothetical/global homology"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
predicted by Codon usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=1
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CADI4850.1"
/db_xref="gi:17428162"
/translation="MNELPLTYAARGVQGLAARPAQLGGLSPDRPMRTWQKKPE
LIRALSPEDMRALVPLSPDPAIKLARREVESRLAQTRGRTFMHGPENRPLPS
RKANRLTLVGVNLVEPAVDALMQRPRTIPDALDDVMISFATDGGGVGHFDSYDV
FLQAGGRWRIRISSNDLTLPDLPKILANFPEBEFLIEPDMLYLPHVADGQV
AEGQGVSVGSRPSRYRELGHGLPGLSOTLENPPEGRYTPDEQKVPRLPEGLPLP
AMVAAALQKNAALRWTELVGEFRGAILTEKQHVETVQRLSLAFTAPARREGIV
LDATQALYDQRFQWINDTFEESDTLLAMLSLSDRGASAEAVDAADLPDMLDT
HAWVEQMLQLAGGAH"
1573. .2097
/gene="slyD"
/note="RSc1149; RS04602"
1573. .2097
/gene="slyD"
/BC_number="5.2.1.8"
/function="cell processes; chaperoning"
/note="Product confidence : probable
Gene name confidence : probable
predicted by Codon usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=1
/product="PROBABLE PKBP-TYPE PEPTIDYL-PROLYL CIS-TRANS
ISOMERASE (ROTAMASE) PROTEIN"
/protein_id="CADI4851.1"
/db_xref="gi:17428163"
/translation="MKIAKNTVSVAVAKLSDAQNVLEESDEPMVYLHGVDGTFPKK
EALDHGQGFETKLEPDDAEGEVHGLKIEPRPRFPELVGMQFEGVPGDD
EDALITYTVDAEDKXVLDGNHPLAGNALRFLRVTVREKTAIEIHGHVGENGLE
VADDEDEPRTTH"
complement(2152. .3216)
/gene="RSc1150"
/note="RS04601"
complement(2152. .3216)
/gene="RSc1150"
/function="miscellaneous; unknown"
/note="Product confidence : probable
Gene name confidence : hypothetical
predicted by Codon usage
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=1
/product="PROBABLE TRANSMEMBRANE PROTEIN"
/protein_id="CADI4852.1"
/db_xref="gi:17428164"
/translation="MARPGAFRAWRGITAAALMLAATAAPAPKGHATAAPAPCF
TFAVGNPQRFEDVAPARALDLIDTEHPAFVHLGNLGRDSCSTDNLLSEEDLTL

Oy		203	ProValGlyAthrAnpProValAlArgyRgheGlyThrAlathrValAlaglySer	222
Db	71124	CCC-----GCCATGGCCAGGTATCGCCCTTC-----GAC	7115	
Oy		223	ThrValThrsEranGlyMetTrpPheSerGlyArgAspGlyAspLeuIleAsnAlaser	242
Db	71157	GACAAGGCACAAAGGCGCATCAGATCGCGCAAGCGGGCGAGCGCGTACGGCTGCC	7121	
Oy		243	AsnAlacGlyThrValIIleGlnAlaAspHisAsnMetsp--GlyAlaserIleValIle	261
Db	71217	GATACCGGAAGAAGATCCACGTGGACCCTTGCGTGGGTACGGGAATCTGTTCATATC	7127	
Oy		262	GlnIsthrThraenglyPheValSerSerTryIleHisIleLysAspAlaGlnValIleThr	281
Db	71277	AAGCACACACGATACCTTCTGACCGCTTACCGGCAACACATTAAGTTCTGGTACCGAG	7133	
Oy		282	GlyAspThrValArgrHrglyGlnArgIleAlaserMetLysAsnGlnProSerGlyAla	301
Db	71337	CAATCCACCGTCAACGAAGAAGGCGCAAGATGCCAATGGCCACTACCGATCGCGACCG	7139	
Oy		302	AlaleudpethupharGlyIleSerArgAsnGlyValTyrrAlaProLeuThrValIleu	321
Db	71397	GTGAACCTACTTGTGAAGTGCAGCGCCGCAACGATTAAGCTCGATCCGATGCGTTCTT	7145	
RESULT	23			
LOCUS	U32753	11435 bp	DNA	linear BCT 29-MAY-1998
DEFINITION	Haemophilus influenzae Rd section 68 of 163 of the complete genome.			
ACCESSION	U32753 L42023			
VERSION	U32753.1 GI:1573701			
KEYWORDS				
SOURCE ORGANISM	Haemophilus influenzae Rd. Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae; Haemophilus.			
REFERENCE AUTHORS	1 (bases 1 to 11435) Fleischmann,R.D., Adams,M.D., White,O., Clayton,R.A., Kirkness,E.F., Kerlavage,A.R., Bolte,C.J., Tomb,J., Dougherty,B.A., Merrick,J.M., McKenney,K., Sutton,G.G., FitzHugh,W., Fields,C.A., Gocayne,J.D., Scott,J.D., Shriver,R., Liu,L.I., Glodok,A., Kelley,J.M., Weidman,J.F., Phillips,C.A., Spriggs,T., Hedblom,E., Cotton,M.D., Uterback,T., Hanna,M.C., Nguyen,D.T., Saudek,D.M., Brandon,R.C., Fine,L.B., Fritchman,J.L., Fuhrman,J.L., Georgagen,N.S., Gnehm,C.L., McDonald,L.A., Small,K.V., Frazer,C.M., Smith,H.O. and Venter,J.C. Whole-genome random sequencing and assembly of Haemophilus influenzae Rd			
JOURNAL MEDLINE	Science 269 (5223), 496-512 (1995)			
PUBMED	95350630			
TITLE	2 (bases 1 to 11435) Tatusov,R.L., Mushegian,A.R., Bork,P., Brown,N.P., Hayes,W.S., Borodovsky,M., Rudd,K.E. and Koonin,E.V. Metabolism and evolution of Haemophilus influenzae deduced from a whole-genome comparison with Escherichia coli			
REFERENCE AUTHORS	Curr. Biol. 6 (3), 279-291 (1996)			
JOURNAL MEDLINE	96398784			
PUBMED	8805245			
TITLE	3 (bases 1 to 11435) White,O., Clayton,R.A., Kerlavage,A.R. and Fleischmann,R.D. Direct Submission			
JOURNAL	Submitted (25-JUL-1995) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA			
REFERENCE AUTHORS	4 (bases 1 to 11435) White,O., Clayton,R.A., Kerlavage,A.R. and Fleischmann,R.D. Direct Submission			
JOURNAL	Submitted (27-SEP-1997) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA			
REFERENCE AUTHORS	The H. influenzae sequence has been updated by R. Fleischmann. New database matches have been assigned. Product names have been improved, and a number of frame shifts have been corrected. We gratefully acknowledge the work of Tatusov et. al. We have incorporated their annotation into the /notes fields of the			
REMARK				

corresponding H. influenzae genes 5 (bases 1 to 11435) White, O., Clayton, R.A., Kerlavage, A.R., Fleischmann, R.D., Peterson, J., Hickey, E., Dodson, R. and Gwinn, M.

JOURNAL Direct Submission
Submitted (28-MAY-1998) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
The whole genome was shifted by 588 nucleotides for a new start on Sep 30, 1996 this sequence version replaced gi:12121405.

REMARK COMMENT
FEATURES
source
1. 11435
/organism="Haemophilus influenzae Rd"
/db_xref="taxon:71421"
complement (106..678)
/gene="HI0699"
complement (106..678)
/gene="HI0699"
/note="similar to GB:LI3261 SP:P30856 GB:Z21496 PID:394720 PID:475995 percent identity: 68.25; identified by sequence similarity; putative"
/codon_start=1
/transl_table=1
/product="peptidyl-prolyl cis-trans isomerase, FkBP-type (sluYD)"
/protein_id="AAC22358.1"
/db_xref="GI:1573702"
/translation="MKRKNNVVISISQVRTQDGLVDEAPANQPLEYLOGHNNLVIGLEKALBEKEVGDFEVVQPEEGYGAISEMMQVRPKDVFQGVDELGVKRPPLADTDIGPVLVTEIDGDEVVVDGNHLAGOELHPTVEVARAEATLBEINAGHVGAHSHDDDEGHGCGCGGHHENHNHNHSCGCGG"
complement (758..1219)
/gene="HI0700"
complement (758..1219)
/gene="HI0700"
/note="similar to SP:P37163 GB:U14003 PID:537096 GB:U00096 PID:1790704 percent identity: 54.20; identified by sequence similarity; putative"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AAC22359.1"
/db_xref="GI:1573703"
/translation="WSKLAEQLQAEETRIIEDLINDSEPNALYIIEHHIAHDFDLLEKIVADFKAGYVSEAESEFPDDDKPFCFDIISEVELKAEIIDAOQKELIPLEKKNGIYDGMGYFEDENADDEGDDGEFLDEDEYCDGDEFFDEDEEPRVH"
1328..2347
/gene="HI0701"
1328..2347
/gene="HI0701"
/note="similar to GB:L07942 SP:P36664 PID:1036739 PID:1036738 GB:U00096 percent identity: 44.24; identified by sequence similarity; putative"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AAC22360.1"
/db_xref="GI:1573704"
/translation="MLEOLPYALAKTPKPTTALKACADPIYKEHLGYEMSGDGEFVAYVKTDCDTLFGVEKLAKFAGVSRNMGYAGLKKRRATVTEGFCLQNEGMETPPDSQFELDGELLVTVRNHRKIRKTSGLSEGYFDILIRGAESSELARLDFVANFGPNQTEORFGRDGNILQALRWAGELIKVDRKRKRSYLSAAREEINLVAAIEKSTINQVLNNDIYVLGSHSWFKADEKEDLTALQVLEMODILLTPALIGDIIILAASEIENEIVNOSADPDLPMKQSRMKAARRPILMKAKGFSWAPEPGLRUKFYLPAGSVATALVRELVINTEE"

gene 2357..3106
/gene="HI0702"
2357..3106
/gene="HI0702"
/note="similar to GB:U00096 PID:1789101 PID:882637 percent identity: 56.50; identified by sequence similarity; putative"
/codon_start=1
/transl_table=1

gene 2357..3106
/gene="HI0702"
2357..3106
/gene="HI0702"
/note="similar to GB:U00096 PID:1789101 PID:882637 percent identity: 56.50; identified by sequence similarity; putative"
/codon_start=1
/transl_table=1

```
/product="stationary-phase survival protein (sure)"
/protein_id="AAC22361.1"
/db_xref="GI:1212202"
/translation="MRILVNDGDFHAGIOLATELRKIAEVIIVAPDNRSAASS
LITVEIPRPHLDNGDYCVNGTPADCVHLNGLSGOVLVSGIAGNMGDDTTY
SGTLAALBERHLDGLPAIVASLDROHETARVVCCLIPKLOHQLNPRRIININP
DLPEELKGYKVCRIIGYRSSSVETIKORDPRDETIVIGPSALPEDESEGDPAVAVN
GVYITPFOADLTLYHSLSLQNWLDQEFK"
3115 3693
/gene="H10703"
3115 3693
/gene="H10703"
3115 3693
/note="similar to GB:L10653 SP:P36684 PID:148929 percent
identity: 72.25; identified by sequence similarity;
putative"
/codon_start=1
/translation="lipoprotein B (lppB)"
/protein_id="AAC22362.1"
/db_xref="GI:1573705"
/translation="MKIFGTMYDKTMEWSKRFAPWLSFVSEIATFPIPPVMLI
PMSMSKRSKAVKFAFYTAIASVIGIIGVATGATMVENIVQOQYAAHMAKAVGM
FEOMGVLVFVAGSPPIPYKVFILCAGVMQMAFPFVITAFVSRILAFILAKLAAMG
GEKFAALRKRSIEIIGMSVVAVIAVFIKLN"
3702 3887
/gene="H10704"
3702 3887
/gene="H10704"
3702 3887
/note="hypothetical protein; identified by Genemark;
putative"
/codon_start=1
/translation="influenzae predicted coding region H10704"
/protein_id="AAC22367.1"
/db_xref="GI:1573706"
/translation="MKLSLVLVAGFLVGCASVDPDALPNCINQVPEGTAVAG
GSFPEIQKNTIPQMK"
3911 5128
/gene="H10706"
3911 5128
/gene="H10706"
3911 5128
/note="similar to GB:L07869 SP:P33648 PID:404099
PID:433185 PID:882635 percent identity: 48.58; identified
by sequence similarity; putative"
/codon_start=1
/translation="lipoprotein"
/protein_id="AAC22363.1"
/db_xref="GI:1573707"
/translation="MKKSFLPLPLSLVLISACTSNPAPISADADGNFSPVIOGVS
NCGAMQPEIQKSLPTTGMNVTPOQNFOPNPTAPACQAFQSPPTVSAFT
VQRTVTRVADVCDOHINIPRNPNTPYKISKSYGNTYKKNKGIMTLLIY
LAGIDVELAALNLSSEPNLSIGOVKISCDIKVTATTVSVQVNSTAIPNV
AVTTPGANGTQISDGTITIGTSEAGTSPSPVATSTQVTSVANNANSTNSV
VAPTLVVMWQMTPSNIGIOGFSSTGNGKIDISGRGCAVKRAAARIVYANAR
GVNGLTIKNDDELISAVAHNDKILVADQCEVKAGQDIARNGSSGTTVTKHPFIRK
KXSDVPRYLPRH"
5261 7846
/gene="H10707"
5261 7846
/gene="H10707"
5261 7846
/note="similar to GB:M64730 SP:P23909 PID:146906
PID:882626 GB:U00096 percent identity: 70.99; identified
by sequence similarity; putative"
/codon_start=1
/translation="DNA mismatch repair protein (mutS)"
/protein_id="AAC22364.1"
/db_xref="GI:1573708"
/translation="MISEENFOQHTPMQOYLKKAENPDILFYRMGDFEYELPYDA
KKAALDLISLTKRGSGAGPIPMAGMRYHAEVGLAKLVOLGSSVALCEIGIPPTS
KGPVERQIVITPGTVSEALPERDNLIAAYOEKELGLTILMTSIRFOLCEP
ADKETRAEIORIAPVELLYCEEFENMAAIEHCKGLRRRRPIWEEFELSTAITILNRGC
```

```
Alignment Scores:
Prid. No.: 0 00115
Score: 219.50
Percent Similarity: 44.69%
Best Local Similarity: 27.84%
Query Match: 13.14%
DB: 1
Gaps: 10

US-10-018-706-2 (1-322) x U32753 (1-11435)

QY 71 AsnArgTgInValysGInGlyAspThrValSerIysIleAlaGlnArgTyrGlyLeu 90
Db 4343 AATACCTATTAAGTAAGTAACAAAGCGGATACATGTTCTTAATCGCTTGGCGGGGAT 4402
QY 91 AsnTrpArgGluIleGlyHisIleAsnAsnLeuAsnSerSerTyrThrIleTyrThrGly 110
Db 4403 GATGTAAAGAAATTGGCGCGGCTGAATACCTATCCGAACTTAATTAAGTTTAGGG 4462
QY 111 GlnTrpLeuThrIleTyrSerGlyAspLeuIleValArgGluIleGlnSerSerGly 130
Db 4463 CAAGTTTAAAGATTCTTAATGTCATTAATAACGAGTTACTCAACGCTTCT 4516
QY 131 ValAsnThrAlaHisThrProSerProValAlaValGlnSerSerArgProProValGln 150
Db 4517 -----GTAACAAACCTCGCAGTCACAGCATCAACAGCAAGCCAGTTAA 4561
QY 151 GlnHisProAlaValGlnIleProThrPro 160
Db 4562 -----CGAGAGTCACTTAT--ACGCCGGCGCAACCGTACTCAATAGATCTGAC 4612
QY 161 -----ProValValValIleValIleValIleValIleValIleValIleVal 175
Db 4613 GGATCATTAATTTGGCCCAATTAATCAAGAGCGTGGACACATCGCTGATCGTGGCA 4672
QY 176 GlnGlnProAlaProValAlaProProValThrGlnAlaProHealThrGlySerSer 195
Db 4673 ACTTCATCAACACAGGTTACATCATCTGTAATTAATGCAATAGCACACCAATTAATTC 4732
QY 196 GlnValMetGlnPheArgTyrProValGlyAlaThrAsnProValValArgArgPheGly 215
Db 4733 AATGTCGTTGCG-----CCAAATC--GCCTCAAT--GTTGTGTGGCAATGCGCG 4777
QY 216 ThrAla-----ThrValAlaGlySerThrValThr-----SerAsnGlyMetTyr 230
Db 4778 ACTTCAGTAAATATCATCAACAGGTTCTCAAGCACATAGCGGTAAACAAAGGAATGAT 4837
QY 231 PheSerGlyTrpArgAspGlyAspLeuIleAsnAlaSerAsnIleGlyThrValIleGlnAla 250
Db 4838 ATTAGTGTTCAACAGACCAAGCTGAAGACGGCTGACACAGGCGCAATAGTATGCA 4897
QY 251 AsnHisAsnMetAspGlyAlaSer-----IleValIleGlnHisThrAsnGlyPheVal 268
Db 4898 GGCATGCTTACGTTGTTACGGTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4957
QY 269 SerSerTyrIleHisIleIleValAspAlaGlnValIleThrGlyAspThrValArgThrGly 288
Db 4958 AGTGCCATGCGCATTAACGACAAATTTCTTGTCGCCGATCAACAGAGTCACAGCGAGT 5017
QY 289 GlnArgIleAlaSerMetIleAsnGlnProSerGlyAlaAlaIleuPheGlnPheArgTyr 308
Db 5018 CAACACATCCCAAAAAGGGATGCTGTGTAACAAATCCGTGAACATCCACTTGAATTT 5077
QY 309 SerArgAsnGlyValTyrValAspProLeuThrValIleu 321
Db 5078 CGTATTAAGTAAATCATGAGATCCAGTACGTACCTA 5116
```


RESULT 24
AE011804
LOCUS
DEFINITION
Xanthomonas axonopodis pv. citri str. 306, linear BCT 29-MAY-2002
the complete genome.
ACCESSION
AE011804 AE008923
VERSION
AE011804.1 GI:21107911
KEYWORDS
Xanthomonas axonopodis pv. citri str. 306.
Xanthomonas axonopodis pv. citri str. 306.
Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
Xanthomonas.
REFERENCE
AUTHORS
1 (bases 1 to 11555)
da Silva, A.C.R., Ferro, J.A., Reinach, F.C., Farah, C.S., Furlan, L.R.,
Quaggio, R.B., Monteiro-Vitorello, C.B., Van Sluys, M.A., Almeida
Utr., N.F., Alves, L.M.C., do Amaral, A.M., Bertolini, M.C.,
Camargo, L.E.A., Camarotte, G., Cannavan, F., Cardoso, J.,
Chambergo, F., Clapina, L.P., Ciccarelli, R.M.B., Coutinho, L.L.,
Cursino-Santos, J.R., El-Dorry, H., Faria, J.B., Ferreira, A.J.S.,
Ferreira, R.C.C., Ferro, M.I.T., Formighieri, E.F., Franco, M.C.,
Greggio, C.C., Gruber, A., Katsuyama, A.M., Kishi, L.T., Leite
Utr., R.P., Lemos, E.G.M., Lemos, M.V.F., Locali, E.C., Machado, M.A.,
Medeira, A.M.B.N., Martinez-Rossi, N.M., Martins, E.C., Meidanis, J.,
Menck, C.F.M., Miyaki, C.Y., Moon, D.H., Moreira, L.M., Novo, M.T.M.,
Oura, V.R., Oliveira, M.C., Oliveira, V.R., Pereira Jr., H.A.,
Rossi, A., Sana, J.A.D., Silva, C., de Souza, R.F., Spínola, L.A.F.,
Taktai, M.A., Tamura, R.E., Teixeira, F.C., Tezza, R.I.D., Trindade dos
Santos, M., Truffi, D., Tsai, S.M., White, F.F., Setubal, J.C. and
Kittajima, J.P.
TITLE
Comparison of the genomes of two Xanthomonas pathogens with
differing host specificities
JOURNAL
MEDLINE
Nature 417 (6887), 459-463 (2002)
PUBMED
22022145
12024217
2 (bases 1 to 11555)
da Silva, A.C.R., Ferro, J.A., Reinach, F.C., Farah, C.S., Furlan, L.R.,
Quaggio, R.B., Monteiro-Vitorello, C.B., Van Sluys, M.A., Almeida
Utr., N.F., Alves, L.M.C., do Amaral, A.M., Bertolini, M.C.,
Camargo, L.E.A., Camarotte, G., Cannavan, F., Cardoso, J.,
Chambergo, F., Clapina, L.P., Ciccarelli, R.M.B., Coutinho, L.L.,
Cursino-Santos, J.R., El-Dorry, H., Faria, J.B., Ferreira, A.J.S.,
Ferreira, R.C.C., Ferro, M.I.T., Formighieri, E.F., Franco, M.C.,
Greggio, C.C., Gruber, A., Katsuyama, A.M., Kishi, L.T., Leite
Utr., R.P., Lemos, E.G.M., Lemos, M.V.F., Locali, E.C., Machado, M.A.,
Medeira, A.M.B.N., Martinez-Rossi, N.M., Martins, E.C., Meidanis, J.,
Menck, C.F.M., Miyaki, C.Y., Moon, D.H., Moreira, L.M., Novo, M.T.M.,
Oura, V.R., Oliveira, M.C., Oliveira, V.R., Pereira Jr., H.A.,
Rossi, A., Sana, J.A.D., Silva, C., de Souza, R.F., Spínola, L.A.F.,
Taktai, M.A., Tamura, R.E., Teixeira, F.C., Tezza, R.I.D., Trindade dos
Santos, M., Truffi, D., Tsai, S.M., White, F.F., Setubal, J.C. and
Kittajima, J.P.
TITLE
Direct Submission
JOURNAL
Submitted (28-NOV-2001) Departamento de Biologia, Universidade de
Sao Paulo, Av. Prof. Lineu Prestes 748, Sao Paulo, SP 05508-900,
Brazil
FEATURES
source
1. 11555 location/Qualifiers
/organism="Xanthomonas axonopodis pv. citri str. 306"
/strain="306"
/db_xref="taxon:190486"
/note="pathovar: citri"
186. 1850
/gene="pyrG"
/note="XAC1716"
186. 1850
/gene="pyrG"
/note="XAC1716"
186. 1850
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=1
/product="CTP synthetase"
/protein_id="AA036583.1"

gene
CDS
/db_xref="GI:21107912"
/translation="MTPLIFVTGGVSLGKGIAAASLAIIEARGKVTMMKLPYI
NVDGMSPROHGVVYTDGAEITDLGLGKRIAVRRLSRKNSVTRGRIYENTYRRE
RGDYLGAATQVPIHITDEIRRCIDENADAGEADVILIGTVGPIEGLPELAEIQR
TERGAEKAMPHMLTLVPIYIAAGELTKRTPQHSIKELRSITIQDVLICRSEQAPDS
ERRRIALFTNVSEBAVISCDDIDVLYCMPLELHQGLDELVTDQFKLRDRAADLS
MAAVDAVKHPLDEVNIAVGVKVDHDAKYSVAEALRHGRLRQRTKNLMAEADL
EGSDMSALODIDGLIVPGFGDRGFEQVTSKPAROKYVFGICYGMQAAVDAAR
HVAIDLAASTENDROS PHPVIGLITEMRTATGVEKRDKSDGLGLOSRK
PGTLAEVYKGVDAERHRRHRYEPNNVYRQLPDAGIVSGKSDPLVEVELPRDT
HPWFLACAPHEFLSTPRDHPFLTGVRAREKAGKLLKERA"
2202. 3032
/gene="kdsA"
/note="XAC1717"
2202. 3032
/gene="kdsA"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=1
/product="2-dehydro-3-deoxyphosphoocronate aldolase"
/protein_id="AA036584.1"
/db_xref="GI:21107913"
/translation="MKLCDFEVLGDQPLFIAGPCVIESMQLQDVAKLTKEITGKL
VNFIFKSPFKANRSTGTSFRGPIELKVLDAVKQIGVPLTDVHAYPNMVA
VVDVLOTPEALVQITDFIKVCAAGKPVNIKQGFLLPMDPKPVDAKSTGNOIV
CBGASFGYNNIVSDMSLSVMRTGCPVDFATHSVOLPGCGSSSGGOREFVLA
RAAVAGTISGLPAETHRDPKSLSDGNANPLDMELLLETIMELDTVTKHGAPRA
"
3138. 3389
/gene="XAC1718"
3138. 3389
/gene="XAC1718"
/note="putative; ORF located using Glimmer/Genemark"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AA036585.1"
/db_xref="GI:21107914"
/translation="MAVMVILTRSTGPPGDBAMQTRLPKRTITLAVMAAVGSL
ATGSAALRLRQPEVALAULTLVLVLLSLGLMA"
3514. 4806
/gene="eno"
/note="XAC1719"
3514. 4806
/gene="eno"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=1
/product="enolase"
/protein_id="AA036586.1"
/db_xref="GI:21107915"
/translation="MTTIATILAREILDSGNPTLEAVTLDDSGFRAAVPSGASTG
TKAEVIEDGDGKTYRIGKRVAVDNNVGIATGLKADQGLDGLDGTEN
KGRGANALGVSLAAHVAASAKPOLVYLSITADVALPVMNNIINGADAN
NDROEFMTLVGCSFSFSAIRAGTEIFSLKSYLKHGSLTAGDGGGAPDPRSRY
BALDTIIEATGKAGTYRAGEDILUGDVIASSEFYDNGKYNLGVENKRLTSEFVPLAD
VWAQYPLISIEDGLAEDMWGKLLTVRVEKQVLDVDDLFTVNPKLFKQSIDSGTAN
AIIKKNQIGTLTETLEIAIMAHANAYASIVSHSGSTETTTIADIVATTATQIKTG
SLGSDRVAKYNQLRIEQLASGARVAGDAPFVIRK"
4810. 5175
/gene="XAC1720"
4810. 5175
/gene="XAC1720"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AA036587.1"
/db_xref="GI:21107916"
/translation="MRWRWLLLVLAVLAMLQYRFWFGPNSGSEVMLEAQAHQIQ

TITLE Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1
JOURNAL Science 286 (5444), 1571-1577 (1999)
MEDLINE 20036896
PUBMED 10567266
REFERENCE 2 (bases 1 to 14193)
AUTHORS White, O., Eisen, J. A., Heidelberg, J. F., Hickey, E. K., Peterson, J. D., Dodson, R. J., Haft, D. H., Gwinn, M. L., Nelson, W. C., Richardson, D. L., Mofatt, K. S., Qin, H., Jiang, L., Pamphile, W., Crosby, M., Shen, M., Vamathevan, J. J., Lam, P., McDonald, I., Utterback, T., Zalewski, C., Makarova, K. S., Aravind, L., Daly, M. J., Minton, K. W., Fleischmann, R. D., Ketchum, K. A., Nelson, K. E., Salzberg, S., Smith, H. O., Venter, J. C. and Fraser, C. M.
TITLE Submitted (08-NOV-1999) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
JOURNAL
FEATURES
Source Location/Qualifiers
1. 14193
/organism="Deinococcus radiodurans"
/strain="R1"
/db_xref="taxon:1299"
/chromosome="1"
/complement (90..1598)
/gene="DR2286"
/complement (90..1598)
/gene="DR2286"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AA011844.1"
/db_xref="GI:6460106"
/translation="MCCDNVPRSPHSMTARVNPTRYLITRSCIEGSMRLKFNRA SPESGVPQVDDRGKRYPAVDNRQWVGIGELYDHLGVNDUTLTPYTRGKRA VETVYKRYSTPRARQTRKQKPPSSGRTETTRREBGESEASRQASPSRREYRQQA GSILGRTPASGGSNLSGLGEERRPATGEEGEPSPSLKPIAPFTQVQPOGQELT LPASLASFGTRSEABEVPAPGAESGSPDPVALSPGSPVPAQVEVSDAPAR AQGLAEQIVPERLASGRLTLRGVRLTADLGPAYGVTLATDRALSSPEDR SGDDRLITTEEDRARGVRLTREALVSLIENARLAPLGLDLRGYRAGDDVLD SA ASVAELVSAHLAQRGTFTYLLTISQAPASLYNVSLAQLNGVSTDDMNLIDTL ARPPLAVTLPENQPLNRSTVLDLADLVAGVRRRLRAPGGTGGPRG"
/gene="DR2287"
1808..2098
/gene="DR2287"
/note="similar to GB:AB000782 percent identity: 64.18; identified by sequence similarity; putative"
/codon_start=1
/transl_table=1
/product="transcriptional regulator, AsnC family"
/protein_id="AA011834.1"
/db_xref="GI:6460096"
/translation="MSGMTAIVWQAEHRRIQETAAALAVSVREVSVTGEMDIV AIVLALEYDLDDVVTGHLKRIKIDIMKTQMLAFRTYNEVLDOGFGLDG"
/complement (2232..3733)
/gene="DR2288"
complement (2232..3733)
/gene="DR2288"
/product="Ribosomal RNA"
/complement (4055..4567)
/gene="DR2288"
complement (4055..4567)
/note="similar to GB:U00096 PID:1787283 percent identity: 64.29; identified by sequence similarity; putative"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AA011835.1"
/db_xref="GI:6460097"
/translation="MPLELVQGDIAHQPDVAVTANKQLMGGGVNDGVTHRAQPR LQAIRPIGTPGTGTAITPAFDLERQGVKVIHAVGPIWGGQHGAEMLLAGVRES LRLVENGCRSVAFPSISTGVYGYPLDRAAIALATLIDFLRSHPDLSVRVNLGADAL

gene HPERALAOI"
complement (4567..5385)
/gene="DR2289"
complement (4567..5385)
/gene="DR2289"
/note="similar to PID:1651722 percent identity: 68.11; identified by sequence similarity; putative"
/codon_start=1
/transl_table=1
/product="guanylate kinase"
/protein_id="AA011836.1"
/db_xref="GI:6460098"
/translation="MGSSQAPGAERGTGTDARTDEAVSVASACMASAPQDRPA GTPQGRGLIVTGASGVGKTLRERWLAGQVFTSWTRPREPVEYNGDYVY SPAEPLAQONGFLEHAGVGNHGYPTPIEALERGGDDVLEIVEGAMQKDM GEOAIIYTPMPSLTBLRRLTGRATETPERIKRLTRADEIQAADPRVYVNDL DRAVSELAVQQAERAAQKAERHTPEEQARALADVNSTALREALQOVES"
/complement (5351..6193)
/gene="DR2290"
complement (5351..6193)
/gene="DR2290"
/note="similar to GB:A009126 percent identity: 55.74; identified by sequence similarity; putative"
/codon_start=1
/transl_table=1
/product="pseudouridylylate synthase I"
/protein_id="AA011837.1"
/db_xref="GI:6460099"
/translation="MDADPERLYAPPGFTRLRLTVAMGDRDYGAGQORNAVQGT LQALNALGEGALRPVSAGRTDAGVAEMPLHMDVPTFRVPLVOLPPALNMLPA SVSLVSAEVPAPGFARFSCTERRYVRLVWAPQRLMAGRALHYGPIDAEAMNRA AOSLIGLHDPAPATREDROTVDRLRLTEVQPGELDIHAGESFLRHVRGIVGTL LTVGQOKLGAEEVGLASERRARAGANVPGGILYFAGAYGRVREGRGNGEPEPT HRRR"
/gene="DR2291"
6309..7673
/gene="DR2291"
/note="similar to PID:1653709 percent identity: 59.04; identified by sequence similarity; putative"
/codon_start=1
/transl_table=1
/product="cell wall glycol-glycine endopeptidase, putative"
/protein_id="AA011838.1"
/db_xref="GI:6460100"
/translation="MLANVPLSLIRLOMKSSPPAPVLAGVTGILLAGALLGAGAE SYRVKGETLYRLNALNGLSEETVQOANPYLRGHALYACQMLTIPKPIPGTFYR KGBDLKKAQRLVSEGDIRENDPOIDRSGALVAGVLRIPARLIAQRAEOAAQOK AAOQKAAQOQAAQAAQAAKRAAQAQAAKKKPPPTTHRVEIGDFFSVARRYGIN PTLQGRNPRLAQTLNAGVLSVAPRLPAPVAPARRAPVRLVAPRAPASPTPAP TAP TSDFGWRVLDGEREKQGLDVAAPPGTPIVLAASGRVLOAHDETYGQWGTIVYIQHD GMQTRVAHLRSLSVEAGQVROQGERVAGVSTGRVVGPHLFGLYRWMDPHNPLAFYA GGLD"
/gene="DR2292"
7675..7932
/gene="DR2292"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AA011845.1"
/db_xref="GI:6460107"
/translation="METAPHTAMPSTVWVYELNGKAEATLRQMLELPQAGPAGAE ILDSPPAQGLALVASHMSGPLPELTPPAGAKHWTFRVLAQR"
/gene="DR2293"
7988..8956
/gene="DR2293"
/note="similar to GB:D12650 SP:P30865 PID:1208970 PID:216647 GB:U00096 percent identity: 45.69; identified


```
/note="residues 1 to 333 of 333 are 58.65 pct identical to
residues 1 to 379 of 379 from E. coli K12 : B2742;
residues 1 to 333 of 333 are 58.65 pct identical to
residues 1 to 379 of 379 from Genpept :
>gb|AA657849.1|AB005502.3 (AB005502) lipoprotein
[Escherichia coli O157:H7 EDL933]"
/codon_start=1
/transl_table=1
/product="lipoprotein"
/protein_id="AA064418.1"
/db_xref="GI:21957563"
/translacion="WSTGSEPMIRLRRAACTVGLVLMVCTNDNSTAPISVGDGRS
GTMLSKANTDSSGGRIYVRSYDNI PKGSYGANTYTVRGDTLFIAMITGNDPRDLA
AKNIIAPYSLAQSIQIOLGNGSGGMAADATSSGIAQSPSNIQNTTIOHTTV
DOSTYASVNGSKONGKMLPSSGAVVATPATPSSSVSPASNGPVGSRMPT
DKGTIDSPSASEGNGKIDIDAGSGOPIATASRGVAGNALRGYGLTIKINDDY
LSAYAHNDMLVRCQEVKAGKIAIMGTSTSVRLHFIRYKGSVNPRLRYLPOR"
1139. .2137
/gene="rpos"
/note="Y0834"
1139. .2137
/gene="rpos"
/function="enzyme; global regulatory functions"
/note="synthesis of many growth phase related proteins;
residues 1 to 332 of 332 are 92.46 pct identical to
residues 1 to 330 of 330 from E. coli K12 : B2741;
residues 1 to 332 of 332 are 97.28 pct identical to
residues 1 to 331 of 331 from Genpept : >gb|AA63391.1|
(U16152) Rpos [Yersinia enterocolitica]"
/codon_start=1
/transl_table=1
/product="RNA polymerase, sigma S (sigma38) factor"
/protein_id="AA064419.1"
/db_xref="GI:21957564"
/translacion="WSOSTLKNVLEHEDADPDENSTETETDEKALVDDETESLAD
DELAQVTOVLDATOLYLGEIGSPILTEBEYFARRALRGDVSSRRMISNLR
LVVIARYSNRGLADLIDIEGNGLIRAVEKDPGFRSTYATWIRQITISRAI
MNQRTIRLPIHIVELNVYLRTAERLSHKLDEPSAEIABOLDKPVDYRMLRLN
ERITSVTPPLGGDEKALDILSDENENGPEDTODDMKOSIQVWMLPELAKAREV
ARRGLGAEAAITVEDVREIGLTRERVRQIQVEGRRLREILQGISIALRRE"
complement (2497. .5052)
/gene="mts"
/note="Y0835"
complement (2497. .5052)
/gene="mts"
/function="enzyme; DNA - replication, repair,
restriction/modification"
/note="residues 5 to 851 of 851 are 84.09 pct identical to
residues 5 to 853 of 853 from E. coli K12 : B2733;
residues 5 to 851 of 851 are 84.09 pct identical to
residues 5 to 853 of 853 from Genpept :
>gb|AA657842.1|AB005501.1 (AB005501) methyl-directed
mismatch repair [Escherichia coli O157:H7 EDL933]"
/codon_start=1
/transl_table=1
/product="methyl-directed mismatch repair protein"
/protein_id="AA064420.1"
/db_xref="GI:21957565"
/translacion="WKNNDKUDSHTPMGOYLRLKAGPEILLFRMGDYELFYSDA
KRAQQLDILSLKRGASAGEBIPMAGVPHYSIENVLAVQLGESAACEIGDPATS
KGPYERKVRIVTPGTTISDBALQERQONLAAIWQDAKGFYATLIDSSGFRVAP
ADLEMAELQRTNPDAELLPENPEPMSLIEHRGLRRPLMEPELDIAKOOLNIOF
TRDILIGFEVROHIALRAAGLLOVVKQRTSLPHIRGLTMRERQOGIINDAATRN
LEIYONLSGSENTLAALDSCVTPMGRMKRMIMRIRIVITDQQAIGIGQDI
LAEOQPLRQVQDLERITLRLALTKTARRDILRKHHAQQLPEIRHLQPIDVPHVN
LISQVQFDELQDLLEAIVETPPVLVADGVASGVNALEDWMLADGATIDVLDL
EYDERKGLDGLTKVGFNGVHYIOWRSQSHVPIHYVRQTKNAERYIIBELKE
YEDKYLKSKGALAEKGLYEIIFDLPLPELOLSANALAEPLVAKOOLNIEKTN
YSCPTLSDPKGIKIMGRHVPVEQVIGKEPISNPLTSPORMLITGPMNGKSTIIL
ROTALIVLAILGSGVVPADQATGPIIDITFRVGAADLASGRSTFMVETNPATIH
NATESLIVLMDIEIGKISTYDGLSLAMCAEKLASRIKAMTLPTHIFELTLPKME
GAVVNHDLALHGETIAFMHSVQEGAAKSYGLAAALGAPVDEVIKARQKLELES
LSNNAASTIDSGWTLNBERIPRAVEALEALDPSLSRQALEWIVYRLKNMV"
```

```
5549. .5683
/gene="Y0836"
5549. .5683
/gene="Y0836"
/function="unknown"
/note="residues 1 to 42 of 44 are 39.53 pct identical to
residues 84 to 126 of 698 from Genpept :
>gb|AF28325.1|AF204952_3 (AF204951) Esv-1-144 [Ectocarpus
siliculosus virus]"
/codon_start=1
/transl_table=1
/product="hypothetical"
/protein_id="AA064421.1"
/db_xref="GI:21957566"
/translacion="WAGMKLQELQIRNEGHCPASELKDQKRLLETSSSDCRGVG"
complement (5846. .6301)
/gene="rpiB"
/note="Y0837"
complement (5846. .6301)
/gene="rpiB"
/function="enzyme; central intermediary metabolism;
Non-oxidative branch, pentose pathway"
/note="residues 1 to 147 of 151 are 44.21 pct identical to
residues 1 to 146 of 149 from E. coli K12 : B4090;
residues 2 to 150 of 151 are 54.36 pct identical to
residues 1 to 148 of 151 from Genpept : >gb|AA03729.1|
(AB006201) RpiA [Pasteurella multocida]"
/codon_start=1
/transl_table=1
/product="ribose 5-phosphate isomerase B"
/protein_id="AA064422.1"
/db_xref="GI:21957567"
/translacion="WLSIATGCDMAIEMNQIYFILOSKNITVVDSCDQOERTW
PDIAQVLAIAKAGHERGILICGTGMSISAKVGVRAQCHDIFSACRARSNN
AQIICLARGPEPLATLILEWDAEFDGSGSTEXVRLRYEQSLHD"
6751. .7899
/gene="Y0838"
6751. .7899
/gene="Y0838"
/function="putative"
/note="similar to 2,3-butanediol dehydrogenase; residues
30 to 375 of 382 are 34.00 pct identical to residues 1 to
345 of 349 from Genpept : >dbj|BA060856.1| (AB063194)
acetylacetoin reductase [Bacillus cereus]"
/codon_start=1
/transl_table=1
/product="putative dehydrogenase"
/protein_id="AA064423.1"
/db_xref="GI:21957568"
/translacion="MCDKPLKESMSHENLKGOMAGEIPETMRVAVVAGPRODFE
RVYPTIDAKELIVKVECGICACADTYARBEABSPFKDDKRPATIKAMITGHERIGH
LVGAGAVEGELDRVTSQIIVPCWGRFCHNQYVMWCEHDIYGFQSNVGAABEY
IKFTKEAIVNHVPHLPLEKAILIEPYACAHAVQAKIKIDVVAVLKGSTLIGMI
GAIRKSGAAGCLIVDLSDDDLALAKKGCADIVLNPBDDVNTLLKEMTVGCGCDIYID
ATGQAKAVEOGLTLRLKLGFEVRSVFEDPYTVMSIISDKELDVLSHGIPYCFPL
VIEGISNGDLPTBEVWHILPLEKAYAGIELVKRGVDSLKVVLPNLT"
7963. .8733
/gene="Y0839"
7963. .8733
/gene="Y0839"
/function="putative"
/note="residues 5 to 256 of 256 are 54.76 pct identical to
residues 3 to 254 of 254 from Genpept : >emb|CAC95595.1|
(AU596164) similar to dehydrogenase/reductase [Listeria
innocua]"
/codon_start=1
/transl_table=1
/product="putative dehydrogenase"
/protein_id="AA064424.1"
/db_xref="GI:21957569"
/translacion="WPNDYKWFDTAFSLSGKVAALITGANGTIGHAIABLYAKARV
LMDCADNVAEIABOLDRDNVAGLCHDVSDSQSVQAAVQAIGAFGQDILVNSAGIAA
```

LDPAEKREQDMDRTIDINLKGVFLMCEVSKHFIQHGKIVNLASQGVVALENHL
AYCASKFVGITKVLALFWPLDIQVNAISPTVLTALGKAMSGQLAEDMKIIPAR
RFAYPAEVAICALFLASDAAMITGENIVIDGYTIO"
8732 8831

gene
CDS

```

//function="putative"
//note="residues 30 to 351 of 362 are 63.35 pct identical
to residues 1 to 332 of 330 from Genpept :
>emb|CAC84700.1 (AL603663) putative dihydroxyacetone
kinase protein [Simorhizobium meliloti]"
/codon_start=1
/transl_table=11
product="putative dihydroxyacetone kinase"

```

shikimate kinase AroL or AroM SW:AROL_ERKCH (P10860) (173 aa) fasta scores: E(): 0, 63.7% id in 168 aa"
/codon_start=1
/transl_table=11
/product="shikimate kinase II"
/protein_id="CAC92450.1"
/db_xref="GI:15981151"
/db_xref="SPTREMBL:Q8ZC15"
/translation="MTQTFVNGARGAGKTTIGKALQALGYRPVTDLEFMQTSQNTVAEVESEMDGDFRLRESMALQAVTPATPVATGGAAVLSSENAPFMDHGRVYLLAASAAVLAKRLAEDEEAPRSLTGKPIVEIIDLVLASREALYQDVAAHVHLDQTPSYLVQIOLMLTGSEMWK"
complement(209..709)
/gene="YPO3215"
/note="Pfam match to entry PF01202 SKI, Shikimate kinase, score 245.00, E-value 1.1e-69"
complement(470..544)
/gene="YPO3215"
/note="P801128 Shikimate kinase signature."
complement(668..691)
/gene="YPO3215"
/note="P800017 ATP/GTP-binding site motif A (P-loop)."
complement(962..1519)
/gene="YPO3216"
complement(962..1519)
/gene="YPO3216"
complement(962..1519)
/note="Similar to Rhizobium meliloti phospholipid N-methyltransferase PmtA TR:AA610237 (EMBL:AF201699) (200 aa) fasta scores: E(): 1.1e-07, 25.9% id in 185 aa, and to Bradyrhizobium japonicum PmtA protein TR:Q91CT2 (EMBL:Y09633) (199 aa) fasta scores: E(): 0.0057, 24.2% id in 178 aa, and to Klebsiella pneumoniae hypothetical 22.7 kDa protein TR:O87755 (EMBL:A011907) (196 aa) fasta scores: E(): 0, 52.9% id in 191 aa"
/codon_start=1
/transl_table=11
/product="putative methyltransferase"
/protein_id="CAC92451.1"
/db_xref="GI:15981152"
/db_xref="SPTREMBL:Q8ZC14"
/translation="MNTIVRRNFKSFQSLQRFIAPRTGSLVPSWPMLQAMVLDIOWTONINIAEGADGVLTKRILSHMSNNSIGAEIQHPFVHLDKRSRLQIARPSAEQOLDQVDIVPCCPLPILSTPTKXISITILQOQRLRARGVLTLYQSHLSSELTLSRFNNMKKIRVVRNPFPAIVYICHPN"
1787..2002
/gene="YPO3217"
1787..2002
/gene="YPO3217"
/note="no significant database hits"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAC92452.1"
/db_xref="GI:15981153"
/db_xref="SPTREMBL:Q8ZC13"
/translation="MIGTDVGVIVANVIMAEITNDGRSISCHFFIOSPFAHPLNVDIKNIASPHVLSIVACSIKKAAXKRLS"
complement(2249..2770)
/gene="YPO3218"
complement(2249..2770)
/gene="YPO3218"
/note="Similar to Escherichia coli hypothetical protein YafE SW:YAFB_ECOLI (P30866) (207 aa) fasta scores: E(): 0, 65.1% id in 169 aa, and to Pseudomonas aeruginosa YafE or P3119 TR:O87011 (EMBL:U03274) (187 aa) fasta scores: E(): 0, 55.2% id in 172 aa"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAC92453.1"
/db_xref="GI:15981154"
/db_xref="SPTREMBL:Q8ZC12"
/translation="MQAADKKLTNIEIQGLAESLPFADOSPDIIVTSRYSAHHMDV

-35_signal
 1338..1343
 /note="potential site for katf gene"
 -10_signal
 1359..1364
 /note="potential site for katf gene"
 gene
 1399..1433
 /gene="katf"
 1399..1433
 /gene="katf"
 /codon_start=1
 /transl_table=11
 /protein_id="AA17876.1"
 /db_xref="GI:476826"
 /translation="MSQNTLKVHDL"
 BASE COUNT 381 a 365 c 369 g 318 t
 ORIGIN

Alignment Scores:
 Pred. No.: 0.000263 Length: 1433
 Score: 213.50 Matches: 87
 Percent Similarity: 42.01% Conservative: 55
 Best Local Similarity: 25.74% Mismatches: 137
 Query Match: 12.78% Indels: 60
 DB: 1 Gaps: 12
 US-10-018-706-2 (1-322) x ECOPECMNLPD (1-1433)

QY 21 ILePheGlyValIleThrThyCysIleLeuAlaGlyCysAlaSerLysProThyTyrAsn 40
 Db 382 GTCTACAGCGCAGCAGCAGCAATTCAGCCGCTGACAGCAGCAAAATTCAGGCTACTCA 441
 QY 41 SerTherGlySerGlySerThiArgThySerGlySerGlyGlyLeu-AlaIleGlyse 60
 Db 442 ACAACCCGCAATTCAGCCAGTGCAGCCAGTGCAGCAGCCGATACAGATGAGAAACGG 501
 QY 60 rGlnValIleThrAspSerGln-----GlyValPro-----AsnArg 72
 Db 502 ACCGATGCTGTATACCGTCAATTCAGGACATTCGCCAAAGCAGTTATAGCGCAGTAC 561
 QY 72 gTyTGlnValIleGlnGlyAspThrValSerLysIleAlaGlnArgTyGlyLeuAsnTr 92
 Db 562 CTATACCGTGTAAAAAGCGCACACCTTCTATATGCGCTGATTACTGGCAGACGATTT 621
 QY 92 pArgGlnIleGlyIleIleAsnLeuAsnSerLysThrIleTyThrGlyGlnTr 112
 Db 622 CCGTACCTGCTGTCAGCGCAACAATTCAGGACCAATACCGCTGAACTGTGTCAG-- 679
 QY 112 pLeuThrLeuThrSerGlyAsp-LeuLysValArgGlyArgSerLysSerGlyValAla 132
 Db 680 ----ACCTTGACAGGTGGTATGCTTCCGGTACGCCAA-----T 714
 QY 132 snThrAlaHisThrProSerProValAlaValGlnSerSerArgProProValGlnGln 152
 Db 715 CACTGGCGGAAATCCATTACCC-----AGCCCGACGAGCAGCA 756
 QY 152 tsrProAlaValGlnLysPro-----ThrProProValValValLysLysProThr- 169
 Db 757 AGGA-GTTGTGATCAAGCTGCACAAATTCACCGTTGCTGTGCTGCAGACCGACA 815
 QY 170 -----P 170
 Db 816 TTACGTATTCGAGTCTTCGGGTGACAGAGTGTACAAATGTGCGCAACAACAGC 875
 QY 170 rGlnThrProProValValGlnGlnProValAlaProProValThrGlnAlaProp 190
 Db 876 CAATCTGAGCAGCAGCAGCAGCGCTGTACAGTACCAACAGCAGCAACCGAGCGCA 935
 QY 190 heaIatnGly-----SerSerGlyValMetGlnPheArgTyProValGlyAlat 207
 Db 936 CTGTACAGAGTACATCAACAGTACGCTATCTCCACTGCGCTGCGC-----ACTG 989

QY 207 hrAsnProValValArgArgPheGlyThrAlaThrValAlaGlySerThrValThrSera 227
 Db 990 AGGCGCAAGTATGATGAAACCTTGGC---GCTTCTGAGGGGGC-----AACA 1034
 QY 227 snGlyMetTrpPheSerGlyArgAspGlyAspLeuIleAsnAlaSerAsnAlaGlyThrV 247
 Db 1035 AGGGGATGTATATGACAGGAGCAGCAAGCAGGCAATATTCGACCGCATGATGCCCGC 1094
 QY 247 alIleGlnAlaAspHisAsnMetAspGlyAlaSer-----IleValIleGlnHisThra 265
 Db 1095 TTGTTATGCTGTGATACCGCTGCGCGCTTACCGTATTCGATATCATCAACCAATATG 1154
 QY 265 snGlyPheValSerSerLysThrIleHisIleLysAspAlaGlnValLysThrGlyAspThrV 285
 Db 1155 ATGATTACCTGAGTACGCTGACGCCCATTAACGACACAAATGCTGTGCGGAAACAAGAG 1214
 QY 285 alArgThrGlyGlnArgIleAlaSerMetLysAsnGlnProSerGlyAlaAlaLeuPheG 305
 Db 1215 TTAAAGCGGGGCAAAAATAGCCACATGCTGACACCGGAAACGATTCACACAGCTTGC 1274
 QY 305 luPheArgIleSerArgAsnGlyValTyValAspProLeuThrValLeu 321
 Db 1275 ATTTGAAATTCGTTACAGGGGAATCCGTAACCGCTGCTGTTATTTG 1324
 RESULT 29
 AE005502/c
 LOCUS
 DEFINITION
 Escherichia coli O157:H7 EDL933 genome, contig 3 of 3, section 121
 OF 290.
 ACCESSION
 AE005502 AE005174
 VERSION
 AE005502.1 GI:12517193
 KEYWORDS
 SOURCE
 ORGANISM
 Escherichia coli O157:H7 EDL933.
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.
 REFERENCE
 1 (bases 1 to 10573)
 Perna,N.T., Plunkett,G., III, Burland,V., Mau,B., Glaesner,J.D., Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A., Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L., Grobleck,E.V., Davis,N.W., Lim,A., Dimilante,E., Potamousis,K., Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C., Welch,R.A. and Blattner,F.R.
 Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
 Nature 409 (6819), 529-533 (2001)
 JOURNAL
 MEDLINE
 PUBMED
 21074935
 11206551
 2 (bases 1 to 10573)
 Perna,N.T., Plunkett,G., III, Burland,V., Mau,B., Glaesner,J.D., Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A., Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L., Grobleck,E.V., Davis,N.W., Lim,A., Dimilante,E., Potamousis,K., Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C., Welch,R.A. and Blattner,F.R.
 Direct Submission
 Submitted (22-OCT-2000) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
 TITLE
 JOURNAL
 FEATURES
 source
 1..10573
 /organism="Escherichia coli O157:H7 EDL933"
 /strain="EDL933"
 /serotype="O157:H7"
 /db_xref="taxon:155864"
 /note="enterohemorrhagic"
 misc_feature
 <1..601
 /note="O-island #110; Region of the EDL933 chromosome not homologous to E. coli K-12 MG1655"
 88..528
 /gene="Z4048"
 88..528
 /gene="Z4048"
 /function="putative regulator; Not classified"
 /note="Residues 12 to 146 of 146 are 100.00 pct identical

to residues 1 to 135 of 135 from Genpept 118 :
gi|4887558|emb|CAB4301.1| (AJO06210) transcripional
regulator protein [Escherichia coli]"
/codon_start=1
/transl_table=11
/product="putative regulator"
/protein_id="AAG57847.1"
/db_xref="GI:12517194"
/translation="MIAADSDPGVYMTLRNKAFLQKLEFGQTHARNOHELDPDTKQ
YAVRAIADKRGIVADLEKAVSTKATLMLMRERGVKREHDAADKRRFVNL
TASGEKTLAAIPIGDSVDEFLRLSAEDELFLVLRKMT"
complement (648 . 1640)
/gene="rpos"
/note="Z4049"
complement (648 . 1640)
/gene="rpos"
/function="enzyme; Global regulatory functions"
/note="Residues 1 to 330 of 330 are 99.09 pct identical to
residues 1 to 330 of 330 from Escherichia coli K-12 Strain
MG1655: B2741"
/codon_start=1
/transl_table=11
/product="RNA polymerase, sigma S (sigma38) factor;
synthesis of many growth phase related proteins"
/protein_id="AAG57848.1"
/db_xref="GI:12517195"
/translation="MSQNTLKVHDLNEDAEFDDNRVEVEDEKALVEEPPDNDLAE
LLSGATQRVADATQLYGEIGEPILTAEEVYFARALGDVARSRRMTESNLRLY
VKIARRYGNRGLALDLIEEGNLGLIRAVEKDEPGRFSTVATWMTROTIERALN
OTRTPLPIHTVKEINLYLRPARLSHLDHPSAEELAEQLDKPVDDVSMTLRN
ITSVDTPLGDSERKLLDILADKENGDEPTDDMKOSTVKKLFELNAKOREVLAR
RRLGLGEATLEPVGREICETRRVNOIQVEGLRRLREILQGINLEALFRE"
complement (11703 . 2842)
/gene="nlpD"
/note="Z4050"
complement (11703 . 2842)
/gene="nlpD"
/function="membrane; Macromolecule synthesis,
modification: Lipoprotein"
/note="Residues 1 to 379 of 379 are 99.73 pct identical to
residues 1 to 379 of 379 from Escherichia coli K-12 Strain
MG1655: B2742"
/codon_start=1
/transl_table=11
/product="Lipoprotein"
/protein_id="AAG57849.1"
/db_xref="GI:12517196"
/translation="MSAGSPKFTVRIRIALSLVSLMAGSGDTSNPAPVSSVGNAP
AHTNSGMLITPPPKGTTSTAQPOIQVQPOIQATQPOIQVQPAQPOVOMENG
RIVNROGNIPEKSGSYSGSTYKKGDTLFYIATGTDFDLAQRNNTIOAPYALNVG
RTLOVGNASGRTITGNAITQADAAEGGVVKPAQNSVAVASQMTIYSSSGGSA
NKMLPNKPKATVTGPAVTPVPLASTTEPTVSTSTSPSTVSMWPTBEKVIETGQAE
GNGGIDIGASKGALLATADGAVVYAGNALRGVGNLIIRKNDYLSAVAHNDTMLV
RGEQEVKAGKATMGSTGTSSTRLHPIRRKKSQVNLRLLPQR"
complement (2982 . 3608)
/gene="pcm"
/note="Z4051"
complement (2982 . 3608)
/gene="pcm"
/function="enzyme; Macromolecule synthesis, modification:
Protein - translation and modification"
/note="Residues 1 to 208 of 208 are 100.00 pct identical
to residues 1 to 208 of 208 from Escherichia coli K-12
Strain MG1655: B2743"
/codon_start=1
/transl_table=11
/product="L-isoaspartate protein carboxylmethyltransferase
type II"
/protein_id="AAG57850.1"
/db_xref="GI:12517197"
/translation="MVSRRVOALLDQLRAQGIQDEOVINALAAVPRKFEVDAFEOKA
MDNIALPIGQGITISQPYVMVAMTELLTPOSRLAETIGTSGYOTAILAHVQVCS
VERIKGLQWARRRLKNLDLHNVSTRHGDGQWQARAPDAIIVTAPEIPELMT

OLDEGILVLPVGEHQYLKVRRRGGEFIIDTVEAVRFVPLVKELIA"
complement (3602 . 4363)
/gene="sure"
/note="Z4052"
complement (3602 . 4363)
/gene="sure"
/function="phenotype; Not classified"
/note="Residues 1 to 253 of 253 are 100.00 pct identical
to residues 1 to 253 of 253 from Escherichia coli K-12
Strain MG1655: B2744"
/codon_start=1
/transl_table=11
/product="survival protein"
/protein_id="AAG57851.1"
/db_xref="GI:12517198"
/translation="MRLISNDGCHAPGIQTLAKALREPADVQVYAPDRNRGASNS
LTLESSLRITTFENGDIAYOMGTPTDCVYGVNALMRPRDIVVSGINAGPNDVYI
YSGTVAAMGRLGFPALAVSDGHRHYPAALVSTIIRALCKEPLTKRLININP
PDLPDQIKIRVTRCGTRHPADQVLPQDPKNTLYWISGPGKCDAGGTDPAAVD
EGVSTIPHLVDLTASHADQVSDVMSVGVGTQW"
complement (4344 . 5393)
/gene="ygbO"
/note="Z4053"
complement (4344 . 5393)
/gene="ygbO"
/function="putative enzyme; Not classified"
/note="Residues 1 to 349 of 349 are 99.42 pct identical to
residues 1 to 349 of 349 from Escherichia coli K-12 Strain
MG1655: B2745"
/codon_start=1
/transl_table=11
/product="putative hydrogenase subunit"
/protein_id="AAG57852.1"
/db_xref="GI:12517199"
/translation="MTEFNDLYLHGKPGGTGLKANPEFVVVEDLGFEPDGEHI
LYRLIKNGCNTRPVADLAKFKITHAEVSPAGQKDAVTEQMLCRVPGKEPDS
AFQIRGQVYKARHREKLEIGALKNAFLVLYEVNRDVRDLCDIVKGVNPF
GAGRFQISGNLQCALPMADNTPVYRNRSFSLASASALPNOYAEELKADVNO
VVDADLQGLKSGSWFVATBELVLEORPNDELMTALLPSGEGCTQRELAFCO
AAVAETELQALVREVEYARRMALLYPOOLSNWMDVTVETIRFMLPAGSPATSYV
RELINTGDTAHIAIE"
complement (5390 . 5869)
/gene="ygbB"
/note="Z4054"
complement (5390 . 5869)
/gene="ygbB"
/function="orf; Unknown function"
/note="Residues 1 to 159 of 159 are 100.00 pct identical
to residues 1 to 159 of 159 from Escherichia coli K-12
Strain MG1655: B2746"
/codon_start=1
/transl_table=11
/product="orf; hypothetical protein"
/protein_id="AAG57853.1"
/db_xref="GI:12517200"
/translation="MRIGHGPDVYAFGEGPIIIGVRIPEKGLAHSDGVALLHAL
TDALGAALGDIKGLFPDTPDPAKGDSEELREARRIRIOAKGYTIGANDVTIIAQA
PKMLPHIPOMKVFIAEDLGCHMDVNNKATITTEKLTGTGGEIGACAVALLIKATY"
complement (5869 . 6579)
/gene="ygbP"
/note="Z4055"
complement (5869 . 6579)
/gene="ygbP"
/function="orf; Unknown function"
/note="Residues 1 to 236 of 236 are 99.15 pct identical to
residues 1 to 236 of 236 from Escherichia coli K-12 Strain
MG1655: B2747"
/codon_start=1
/transl_table=11
/product="orf; hypothetical protein"
/protein_id="AAG57854.1"
/db_xref="GI:12517201"
/translation="MATTHLDVCAVVPAGFRRMOTCEPKOYLISGNQTLLEHSVYA

sequence changes. Annotation updates: updated gene identifications and products; all new functional assignments courtesy of Monica Riley; added promoters, protein binding sites, and repeated sequences described in reference 1. The unique numeric identifiers beginning with a lowercase 'b' assigned to each gene (protein- or RNA-encoding) are now designated as gene synonyms instead of labels. This should allow them to be searched for in Entrez as gene names.

FEATURES

source

Location/Qualifiers
1.11457
/organism="Escherichia coli K12"
/strain="K12"
/sub_strain="MG1655"
/db_xref="taxon:83333"
75.1439
/gene="b2740"
75.1439
/gene="b2740"
/function="putative transport; Not classified"
/note="0454; This 454 aa ORF is 34 pct identical (9 gaps) to 444 residues of an approx. 456 aa protein GMP_BACU1 SW: P46832"
/codon_start=1
/transl_table=1
/product="putative transport protein"
/protein_id="AAC75782.1"
/db_xref="GI:1789097"
/translation="MSTITLICALAGVIMLLIVAKVQPFVALLVSLVALAGIPAGEVGKVMVAGMGVIGSVTIIIGAMLMRMIEHSGESLANYSRRLGDKRTI AALTAAPFPGIPVPDPVGGIILAIPIYGVPAKVISPLKGLPAQIMLVHVAAP HGPVAAAGLILHADIIMLTIGIISIPGVGVPAKTIKRGVAMSVEVLEMOA PASREGATKLSDKINPVALSVSLIYVPIALIMAGTASATLMPSPHGLGTLIS PMVIMTAIVTAFWLIALRGMVLOHTSDIMGSLPTPAVILITGAGGVGKTLVS GVGKLANMLOMIDILIPALPISILASQGSATVALITGGILSEAVNGLNPIOC VLTALACFSGSHINDSGFWITKILGSLVADGLKTVLVTITLIGFTGFLITWCY WAVI"
Complement(1533.2525)
/gene="rpos"
/note="b2741"
complement(1533.2525)
/gene="rpos"
/function="enzyme; Global regulatory functions"
/note="f330; 99 pct identical to RPOS_ECOLI SW: P13445; CG Site No. 18208; alternate gene names katF, appR, sigS, otxK, nur"
/codon_start=1
/transl_table=1
/product="RNA polymerase, sigma S (sigma38) factor; synthesis of many growth phase related proteins"
/protein_id="AAC75783.1"
/db_xref="GI:1789098"
/translation="MSQNTLKVDHNEADAFDENGVEVDEKALVEQSPNDIAEEELISOGATRVADATOLYGEIGVSPILTAEEVEYFARALGDVARSRRMESNRLV LKIRYVGNRGALDLIEGNLGIIRAVEKDFPGRFSTVATWVRITIERALNM OTRIPLPIHIVKELNIVYLTARELSHLDEPSEAEIABOLDKVDVDSMLNLER ITSDVPLPGSGSEKALDILADKEKNGEDTODDMKOSIVKMLFELNAROREVLA RPLIGVATLLEVGREIGLTREVRVRIQVGRRLREIITGTGLENIALEFRE
complement(2572.2603)
/note="factor Sigma70; predicted +1 start at 286564"
complement(2588.3727)
/gene="hlpD"
/note="b2742"
complement(2588.3727)
/gene="hlpD"
/function="membrane; Macromolecule synthesis, modification: Lipoprotein"
/note="f379; 100 pct identical to NLPD_ECOLI SW: P33648; CG Site No. 33225; ORFX of D17549, has different start due to frameshift in D17549"
/codon_start=1
/transl_table=1
/product="hlpD protein"
/protein_id="AAC75784.1"

gene

CDS

/db_xref="GI:1789099"
/translation="MSAGSPKFTVRILALSLVSLWLAGSDTSNPPAPVSVGNAP ANTNMILITPPKMGKTSTTAPOIOPVOPOIOIATQPOIOPVOPOIOMENG RIYNNQYNGIPKSGSYGVKQDTELYIMTIGNDPRDLAONNIQAPALANG OTLOVNASGTPIITGNALTOADAQGVVIRKQNSYVVASOPTITYESGEOA NKLPLPNKPAATTVTAPVVPTASTTEPTVYSTSTSTSTMTPEGKVITPEGASE CGKGDIDGSKQALITATDGRVYVAGNALRGYMLIIRKHNDVLSAVAHNDTMLV RQGVYKACQKIXITMGSTGSTRHFEIRKYSNPLKYLPGR"
complement(3867.4493)
/gene="rpm"
/note="b2743"
complement(3867.4493)
/gene="pcm"
/EC_number="2.1.1.77"
/function="enzyme; Proteins - translation and modification"
/note="f208; 100 pct identical to PWT_ECOLI SW: P24206 but includes initiator met; CG Site No. 33221"
/codon_start=1
/transl_table=1
/product="L-isoaspartate protein carboxylmethyltransferase type II"
/protein_id="AAC75785.1"
/db_xref="GI:1789100"
/translation="MVSRRVQALIDQLRAQIODEQVNLAAVPREKFEVDAFEOKA VDNIALPIGQGTISQYVAVRMTLELTPQSRVLEIGTSGVQTAIHLHLYQVS VERRKIGIOMARRRLKDLHNSTHGDQMGQVAPDAIITVAAPETPLATMT QLDGGLIVLPVEEHQYLKRVRRRGGEPIDTVEAARFVPLVKEILA"
complement(4487.5248)
/gene="sure"
/note="b2744"
complement(4487.5248)
/gene="sure"
/function="phenotype; Not classified"
/note="f253; 100 pct identical to PIR: I69732; alternate name ybcC; ORF1 of L07942, has different start due to frameshift in L07942"
/codon_start=1
/transl_table=1
/product="survival protein"
/protein_id="AAC75786.1"
/db_xref="GI:1789101"
/translation="MRLISNDGCHNAPGIOTAKALREPADVOVAPDNRGASNS LTISSLRITFENGDIADVOMGPTDCVYGVNAKRPDIYVSGINAGPNDVYI YSGVAAAMEGRHLGEPALAVSDGHRHYDTAAVTSITRALCKEPLKRLINIV PDLPLDQIKGIRVRCGRHPADQVLPQDPKRNATYIGPPGKCDAGGTDPAAVD EGVYSIRPLHVDLTAHSAQDVSDMNSVGVTQW"
complement(4535.4563)
/gene="sure"
/note="factor Sigma70; predicted +1 start at 2867577"
complement(5229.6278)
/gene="ygbO"
/note="b2745"
complement(5229.6278)
/gene="ygbO"
/function="putative enzyme; Not classified"
/note="f349; was f292 and f255 before splice; This 349 aa ORF is 44 pct identical (8 gaps) to 338 residues of an approx. 344 aa protein y701_HABIN SW: P44039"
/codon_start=1
/transl_table=1
/product="putative hydroxylase subunit"
/protein_id="AAC75787.1"
/db_xref="GI:1789102"
/translation="MIFEDULTYHKGPGSTGLIKNPDPVVEDIGPPDGGHET LVRILKGVNTRFVADLAKFLKHAKEVFPAGKDVADHATBCKLRVKGWPPDIS AFVREGGVAVHARKRLGAKHGAFTLVGVSNRQDVDEKLDIVYKVPYPA GAQRFIGISNLSGAGQWATNTFVRKRNKSTPLSARALRQIVAEKRLKADVNO VVDSDALQLAGRSWFWATTELAELDQRYVNDXELMTTALPSSGEGTORALAFEG

gene

CDS

/db_xref="GI:1789103"
/translation="MSTITLICALAGVIMLLIVAKVQPFVALLVSLVALAGIPAGEVGKVMVAGMGVIGSVTIIIGAMLMRMIEHSGESLANYSRRLGDKRTI AALTAAPFPGIPVPDPVGGIILAIPIYGVPAKVISPLKGLPAQIMLVHVAAP HGPVAAAGLILHADIIMLTIGIISIPGVGVPAKTIKRGVAMSVEVLEMOA PASREGATKLSDKINPVALSVSLIYVPIALIMAGTASATLMPSPHGLGTLIS PMVIMTAIVTAFWLIALRGMVLOHTSDIMGSLPTPAVILITGAGGVGKTLVS GVGKLANMLOMIDILIPALPISILASQGSATVALITGGILSEAVNGLNPIOC VLTALACFSGSHINDSGFWITKILGSLVADGLKTVLVTITLIGFTGFLITWCY WAVI"
Complement(1533.2525)
/gene="rpos"
/note="b2741"
complement(1533.2525)
/gene="rpos"
/function="enzyme; Global regulatory functions"
/note="f330; 99 pct identical to RPOS_ECOLI SW: P13445; CG Site No. 18208; alternate gene names katF, appR, sigS, otxK, nur"
/codon_start=1
/transl_table=1
/product="RNA polymerase, sigma S (sigma38) factor; synthesis of many growth phase related proteins"
/protein_id="AAC75783.1"
/db_xref="GI:1789098"
/translation="MSQNTLKVDHNEADAFDENGVEVDEKALVEQSPNDIAEEELISOGATRVADATOLYGEIGVSPILTAEEVEYFARALGDVARSRRMESNRLV LKIRYVGNRGALDLIEGNLGIIRAVEKDFPGRFSTVATWVRITIERALNM OTRIPLPIHIVKELNIVYLTARELSHLDEPSEAEIABOLDKVDVDSMLNLER ITSDVPLPGSGSEKALDILADKEKNGEDTODDMKOSIVKMLFELNAROREVLA RPLIGVATLLEVGREIGLTREVRVRIQVGRRLREIITGTGLENIALEFRE
complement(2572.2603)
/note="factor Sigma70; predicted +1 start at 286564"
complement(2588.3727)
/gene="hlpD"
/note="b2742"
complement(2588.3727)
/gene="hlpD"
/function="membrane; Macromolecule synthesis, modification: Lipoprotein"
/note="f379; 100 pct identical to NLPD_ECOLI SW: P33648; CG Site No. 33225; ORFX of D17549, has different start due to frameshift in D17549"
/codon_start=1
/transl_table=1
/product="hlpD protein"
/protein_id="AAC75784.1"

promoter

gene

CDS

/db_xref="GI:1789099"
/translation="MSAGSPKFTVRILALSLVSLWLAGSDTSNPPAPVSVGNAP ANTNMILITPPKMGKTSTTAPOIOPVOPOIOIATQPOIOPVOPOIOMENG RIYNNQYNGIPKSGSYGVKQDTELYIMTIGNDPRDLAONNIQAPALANG OTLOVNASGTPIITGNALTOADAQGVVIRKQNSYVVASOPTITYESGEOA NKLPLPNKPAATTVTAPVVPTASTTEPTVYSTSTSTSTMTPEGKVITPEGASE CGKGDIDGSKQALITATDGRVYVAGNALRGYMLIIRKHNDVLSAVAHNDTMLV RQGVYKACQKIXITMGSTGSTRHFEIRKYSNPLKYLPGR"
complement(3867.4493)
/gene="rpm"
/note="b2743"
complement(3867.4493)
/gene="pcm"
/EC_number="2.1.1.77"
/function="enzyme; Proteins - translation and modification"
/note="f208; 100 pct identical to PWT_ECOLI SW: P24206 but includes initiator met; CG Site No. 33221"
/codon_start=1
/transl_table=1
/product="L-isoaspartate protein carboxylmethyltransferase type II"
/protein_id="AAC75785.1"
/db_xref="GI:1789100"
/translation="MVSRRVQALIDQLRAQIODEQVNLAAVPREKFEVDAFEOKA VDNIALPIGQGTISQYVAVRMTLELTPQSRVLEIGTSGVQTAIHLHLYQVS VERRKIGIOMARRRLKDLHNSTHGDQMGQVAPDAIITVAAPETPLATMT QLDGGLIVLPVEEHQYLKRVRRRGGEPIDTVEAARFVPLVKEILA"
complement(4487.5248)
/gene="sure"
/note="b2744"
complement(4487.5248)
/gene="sure"
/function="phenotype; Not classified"
/note="f253; 100 pct identical to PIR: I69732; alternate name ybcC; ORF1 of L07942, has different start due to frameshift in L07942"
/codon_start=1
/transl_table=1
/product="survival protein"
/protein_id="AAC75786.1"
/db_xref="GI:1789101"
/translation="MRLISNDGCHNAPGIOTAKALREPADVOVAPDNRGASNS LTISSLRITFENGDIADVOMGPTDCVYGVNAKRPDIYVSGINAGPNDVYI YSGVAAAMEGRHLGEPALAVSDGHRHYDTAAVTSITRALCKEPLKRLINIV PDLPLDQIKGIRVRCGRHPADQVLPQDPKRNATYIGPPGKCDAGGTDPAAVD EGVYSIRPLHVDLTAHSAQDVSDMNSVGVTQW"
complement(4535.4563)
/gene="sure"
/note="factor Sigma70; predicted +1 start at 2867577"
complement(5229.6278)
/gene="ygbO"
/note="b2745"
complement(5229.6278)
/gene="ygbO"
/function="putative enzyme; Not classified"
/note="f349; was f292 and f255 before splice; This 349 aa ORF is 44 pct identical (8 gaps) to 338 residues of an approx. 344 aa protein y701_HABIN SW: P44039"
/codon_start=1
/transl_table=1
/product="putative hydroxylase subunit"
/protein_id="AAC75787.1"
/db_xref="GI:1789102"
/translation="MIFEDULTYHKGPGSTGLIKNPDPVVEDIGPPDGGHET LVRILKGVNTRFVADLAKFLKHAKEVFPAGKDVADHATBCKLRVKGWPPDIS AFVREGGVAVHARKRLGAKHGAFTLVGVSNRQDVDEKLDIVYKVPYPA GAQRFIGISNLSGAGQWATNTFVRKRNKSTPLSARALRQIVAEKRLKADVNO VVDSDALQLAGRSWFWATTELAELDQRYVNDXELMTTALPSSGEGTORALAFEG

Alignment Scores:
Pred. No.:
Score:

0.0026 Length: 11457
213.50 Matches: 87

Percent Similarity: 42.01%
 Best Local Similarity: 25.74%
 Query Match: 12.78%
 DB: 1
 Gaps: 12

US-10-018-706-2 (1-322) x AE000358 (1-11457)

QY 21 IlePheGlyValIleThrThrCysIleLeuAlaGlyCysAlaSerLeuProThrTyraen 40
 Db 3542 GTCACACGCGAAGCAACCGAATTCACCGCGTGCAGACACCAAAATTCAGGCTACTCA 3483
 QY 41 SerThrSerGlySerGlySerHisThrSerGlySerGlyLeu-AlaIleGlyse 60
 Db 3482 ACAACCGCAATCCAGCGACGACGAGCTGACAGACCGGTACAGATGAAACGCG 3423
 QY 60 rglValIleThrThrAspSerGln-----GlyValPro-----AsnAr 72
 Db 3422 AGCGATCGCTATTAACCGCTCAGTATGGAAACATTCGAAAGGCACTTATACGCGCAGTAC 3363
 QY 72 GTTCGlnValIleGlnGlyAspThrValSerIleAlaGlnAgtTyrglyLeuAsnTr 92
 Db 3362 CTATACCGTGAAGAAAGCGACACCTTTCATATCCGCTGATTCGCAACGATTT 3303
 QY 92 pArgGluIleGlyHisIleAsnLeuAsnSerSerTyrrhIleTyrrhGlyGlnTr 112
 Db 3302 CCGTACCGCTCTCAGCGCAACATATTCAGGACCATACGCGGTGAACGTTGTCAG-- 3245
 QY 112 pleuThrleuThrSerGlyAsp-LeuIysValArgGluArgSerIleSerSerGlyValA 132
 Db 3244 ---ACCTTGACAGTGGGTAAATGCTTCGGTACCGCA-----T 3210
 QY 132 snrThAlaHisThrProSerProValAlaValGlnSerSerArProProValGlnGln 152
 Db 3209 CACTGGCGGAATGCACTTACC-----AGCGCAACGACGACGACCA 3168
 QY 152 isPProAlaValGlnIlePro-----ThrProProValValValIleIysPProThr- 169
 Db 3167 AGCA-GTGTGATCAAGCGCTGCACAAATTCACCGTGTGCTGTCGTCACGACGACA 3109
 QY 170 -----P 170
 Db 3108 TTACGATTCGAGCTTCGGGTGAACAGAGTGTACAAATGTTGCCGAACAAAGC 3049
 QY 170 roThrProProValValGlnGlnProAlaProValAlaProProValThrGluAlaProp 190
 Db 3048 CACTGGCAACGACGCTCACAGCGCTGTACGCTACCAACAGCAACGACGACCGCGCA 2989
 QY 190 heaIaThrGly-----SerSerGlyValMetGlnPheArgTyrrProValGlyAlaT 207
 Db 2988 CTGTACGACAGTACATCAACAGTACGCTATCTCACTGCGGCGTGGCG-----ACTG 2935
 QY 207 hrAsnProValValArgArgPheGlyThrAlaThrValAlaGlySerThrValThrSerA 227
 Db 2934 AGGCGCAAGTATCAAACTTTGGC---GCTTCGAGGGGGC-----AACG 2890
 QY 227 snGlyMetTrpPheSerGlyArgAspGlyAspLeuIleAsnAlaSerAsnAlaGlyThrV 247
 Db 2889 AGGGATTCATATCGCAGCAAGCAAGCAAGCAATTCGACCGCAGATGCGCGCG 2830
 QY 247 alIleGlnAlaSprHisAsnMetAspGlyAlaSer-----IleValIleGlnHisThrA 265
 Db 2829 TTGTTTATGCTGTGAACGCGTGGCGGCTACCGTATCTGATTCATCAACATTAATG 2770
 QY 265 snGlyPheValSerSerTyrrIleHisIleIysAspAlaGlnValIysThrGlyAspThrV 285
 Db 2769 ATGATTACTTACCTGATCCCTACCGCATTAACACAACTGCTGTGTCGGAACAAACGAG 2710
 QY 285 alArgThrGlyGlnArgIleAlaSerMetIysAsnGlnProSerGlyAlaIleAlaLeuPheG 305
 Db 2709 TTAAGCGCGGCAAAATTAAGACCATGCGGTAGACACCGGAACGATTCACACGCTTGC 2650
 QY 305 IupheArgIleSerArgAsnGlyValTyrrValAspProleuThrValIleu 321

Db 2649 ATTTGAAATTCGTTACAGGGGGAATCCGTAACCCGCTGCTATTG 2600
 RESULT 31
 ECUT9579/c
 LOCUS ECUT9579 72221 bp DNA linear BCT 01-JUL-1995
 DEFINITION Escherichia coli K-12 genome; approximately 61 to 62 minutes.
 ACCESSION U29579
 VERSION U29579.1 GI:882596
 KEYWORDS
 SOURCE
 ORGANISM
 Escherichia coli.
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Escherichia coli.
 REFERENCE
 AUTHORS Plunkett, G.P.
 TITLE Direct Submission
 JOURNAL Submitted (20-JUN-1995) Guy P. Plunkett, University of Wisconsin, Laboratory of Genetics, 445 Henry Mall, Madison, WI 53706, USA
 COMMENT This sequence was determined as part of the E. coli Genome Project (Frederick R. Blattner, director) at the University of Wisconsin-Madison. Supported by award HG00301 from the NIH Human Genome Project. The entire sequence was independently determined from E. coli MG1655; overlaps with other sequence determinations are annotated. This entry should be considered somewhat provisional; it will be updated and merged with others at a later date.

FEATURES
 source
 location/Qualifiers
 1..72221
 /organism="Escherichia coli"
 /strain="K-12"
 /sub_strain="MG1655"
 /db_xref="taxon:562"
 /map="approximately 61 to 62 minutes"
 /note="this sequence comprises the following lambda clones: DB609 (EC30K46-3), DB612 (EC19-195), DB615 (EC27-240), DB618 (EC22-172), and DB620 (EC18-1463); MJJannus vector was used for subcloning"
 1..2371
 /note="similar to GenBank Accession Number J02708 (ECOGUT)"
 39..818
 /gene="gucD"
 39..818
 /gene="gucD"
 /EC_number="1.1.1.140"
 /note="CG Site No. 153; alternate gene name srlD"
 /codon_start=1
 /transl_table=11
 /product="glucitol-6-phosphate dehydrogenase (sorbitol-6-phosphate 2-dehydrogenase)"
 /protein_id="AA06924.1"
 /db_xref="GI:882597"
 /translation="MNQVAVITGGGQTGAFLCHGLAAEGYVAVVDIQQDKANVAQ EINAEVESHMAVYGADATSEOSYALASRGDEIFGRVDLVYAGIAKAIFISFOL GDPFRSLQVNLVGYFLCARERSRLIMIDIGRIQLINSKSGKSGKNSGYSAAKFG GVLGTSLADLAIEYGITVYSIMGNLKSFMPSLIPQVATYKLGIRKDYEQYVIDK VPIKRGCDYDVANMLFYASPKASYCTGGSINTVGGYVM"
 572..2371
 /note="similar to GenBank Accession Number X13463 (ECUTWR) 1..1800"
 924..1283
 /gene="gucM"
 924..1283
 /gene="gucM"
 /gene="gucM"
 /function="glucitol operon activator"
 /note="CG Site No. 33198; alternate gene name srlM"
 /codon_start=1
 /transl_table=11
 /protein_id="AA069215.1"
 /db_xref="GI:882598"
 /translation="WVSALITVAVIANCAOLALGKQWISRPNRADFTLCCGGRVGVG SSGRFPKRVVVAALDDQORIVDTLFMKGLTVFARPKIPAITGMHAGDLOPDIYFPH DPLSQNALSLALKLRG"

```

gene      1350..2123
/ gene="gucR"
CDS       1350..2123
/ gene="gucR"
/ function="glucitol operon repressor"
/ note="CG Site No. 152; alternate gene name srlR"
/ codon_start=1
/ trans1_table=11
/ protein_id="AAA69216.1"
/ db_xref="GI:882599"
/ translation="MKPRQAAILEYIQCKSCSVEELAYDPDTTGTTRKIVILE
HAGVIRTYGGVAVLNKESDPIDHKTILINHKKEILAEAVSFTHDSIILIDAGST
VLOWVPLSRFNNTIVMTNSLAIYVALSELNEQITLIPGDTFKKSKSPHQLEAEN
FEHPTFDLFWGIDGIDLNAGVTFENEYTVSKAMCNAREVILMAOSSKFRKSPNV
VCSLESVDKLTIDAGIDPAFRQALEEKIDIVITGESNE"
misc_feature 2121..3054
/ note="similar to GenBank Accession Number X51361
(ECGUT0)"
gene      2155..3081
/ gene="gucO"
CDS       2155..3081
/ gene="gucO"
/ note="alternate gene name srlQ"
/ codon_start=1
/ trans1_table=11
/ protein_id="AAA69217.1"
/ db_xref="GI:882600"
/ translation="MLRQFASRLPERLGDPPVRAANTILHCEGVVVSIGTSGHTG
KRIATLSTGTPAFVVPARALHGDLMISRPVLMPLSYGAKKEIDILIPLEAD
SLIALAMGKSPSGIAAKAVDISVREPCPMPLASTVNTLMGDLAMVMO
ARGREPRASHPGALGARILNHLHMRDDAIPVALTASVMDAMLELRGIG
LVAVDAAQQVQGVPTDDIRRMVCGGALTPPVNEAMTVGGTTLQSSRAIDAKEIL
MKRTIAFVVDENKLTGALNIDPFOGII"
misc_feature 3078..3590
/ note="alternate name ygaA"
/ codon_start=1
/ trans1_table=11
/ label="ORF_f170"
/ protein_id="AAA69218.1"
/ db_xref="GI:882601"
/ translation="MLAGRFADLPHRLSVPLSVPLREGRDDVILLAGYCECCRL
RQGLSRVLSAGARNLQHYSPFGNRELEHRAVVALRATRGSDVILAEQFAR
PEVLTPTREVAAPVVKONLREATEAFQRETRQALQNHNMWAAACAMLETDVANLH
RLAKRLGLKD"
CDS       complement(3610..4668)
/ note="alternate name ygaB; ORF_f352"
/ codon_start=1
/ trans1_table=11
/ protein_id="AAA69219.1"
/ db_xref="GI:882602"
/ translation="MSSONDYQSGRYNIIYSQISEAMSPFVDVLANIAIEIQRGI
GHODRFORLITLRQVLECDASALLRYSRQPIPLAIGLADVLGRPALLEGHRELE
AYARAGDVYRFPADSELPDPYGLI PGQESVHACVGLPLPAGNLGALGLIDMOPG
DOPVPSDEBELILALAAAGLSNMLILEOLESOMGLGDAPREPAVQOTWIGISPG
MTOLKKEITIVASDLNVLISGETNGKEILVKA1HEBASPRVNVPLINCAALRESV
AESRFGHVKGAFTGALISRSKGFPMADNGTLFLDEIGLSIALQAKLRLVLYQYDIO
RVEDRCLRVYRVVIGDP"
CDS       4780..6219
/ note="ORF_c479"
/ codon_start=1
/ trans1_table=11
/ protein_id="AAA69220.1"
/ db_xref="GI:882603"
/ translation="MSIVKNNIHWGQDWEVRDPFHGTGYTLRGSSYSYLIREEK
NLVIDVHKFRFVQNLRNHIDLADIDYIINHAEBDHAGALTELMALQIPDTIYC
TNAIDISINGHHHPMNPNVVKGTDTLDIGKQKQALFVETSMILWPMSPMYLFGDA
VLENDAGQHYCDHLEFNDVDTLPEQCORQYANILFVDTMSNVTPEKITEILGENTL
PADMIATSHGVARNPTQIVELYLKMAADYEDRITTFYDTMSNVTPEKITEILGENTL
AETDPRVAVKFINVASDKNEILTVFKSGVLYGTSWNNWMPRIAGLVEEMTGLR
FANKKASAFSGMSGAVDRLSRLTODGFMGMSLSLKAKMPPDDALDLCREHRELI
AOMKALAPLPQSTVNTVKEETSAITLADIGPMOCSVCOMIYDAKAGPMODVAGT
PSEVVDNPLCECSIGKVFEELASAK"
CDS       6216..7349

```

```

misc_feature 6641..10718
/ note="ORF_c377"
/ codon_start=1
/ trans1_table=11
/ protein_id="AAA69221.1"
/ db_xref="GI:882604"
/ translation="MNSGITVIGSGFAARQLVKNIRKODATPIFLTIIAASMDPEYNK
DLSHVISOGRADLTROTATGERAEOFINHLFPQTVWTDIDAEARVYSQNMOMYDK
LYATGASAFVPPVPGSELMTLINSQOEYACETOLRDAARVLIYGGGILGSELANDP
CPAGKATVITDMSASTLASIMPEVSSRLCHRLNREMGVHLLSLOGLEKTDGLOA
TUDORNYIYDAVIAATGAPETRLARBAITLRNGVCVSVYQTSNTDYLALGDCAE
INCQVLPRIOPILSAMVAKNLGNNTPKLPMILVKITPPLPHLAEETQRDLR
WQITREQCNVAGVDADOLRAFYVSEDMKRAFGILKTLPM"
misc_feature 6641..10718
/ note="similar to GenBank Accession Number D14422
(ECOHYA)"
gene      7477..9729
/ gene="hyda"
CDS       complement(7477..9729)
/ gene="hyda"
/ complement(7477..9729)
/ function="transcriptional regulatory protein"
/ note="CG Site No. 18232; alternate gene name hypF"
/ codon_start=1
/ trans1_table=11
/ protein_id="AAA69222.1"
/ db_xref="GI:882605"
/ translation="MAKNTSCGVQLRIRGVQGVFPFVWQIAQQLNLHGVDVNCDD
GVEVRLDEDETFVLOVHOCPPLARISVEREPFISQULPTEFTIRQSGTGMNTQI
VPDAATCPACIAENKTGERRRYPFINCHGCRFTIIRAMPEDRPFYMAAFPLCP
ACKEVYEDPLDRPHAQVAPCEGCPHEWVSHGEHBOAALQALQALQKMGKIVAI
KGIQGFPLACDARNSNAVITLRARKEHFAKPLAVLEADGLPDAALQTLTPPAPIV
LVKKYVPELCDIAPLNEGVMLPANTLOHLLLOELQCPLEWTSGLSGKPPALIN
EOLADLOGLADGFLINRDIYOMDSVYRSGEMLRSGVYPDALPFGKXNP
PVLCTGADLNKTFVLRGOAVLSQHGDSDDIOQOMBEALRMQNIYDFPQYVY
HDAPRGVCCRYVRECEHIGLPAVALPGGDUALAKQPMNLALCAAPATLSBBAALAE
NGLWGECCRYVRECEHIGLPAVALPGGDUALAKQPMNLALCAAPATLSBBAALAE
TASYOHMWSTVLAIRGINAPLASCGSPFPDAVALCAAPATLSBBAALAE
LASCHGVTRHTVPRVDNOLDLATPQOMLNQAPVNOAMAFHDLAAGPAAIAME
QATIRGTTTLVFSGCVIHNRLRLARLHYLADTLFLPQSLPADGSLSGQGVIAAA
RWLAGEVQNG"
CDS       complement(9882..10409)
/ gene="hyda"
/ complement(9882..10409)
/ function="electron transport from formate to hydrogen"
/ note="CG Site No. 33521"
/ codon_start=1
/ trans1_table=11
/ product="4Fe-4S iron-sulfur protein"
/ protein_id="AAA69223.1"
/ db_xref="GI:882606"
/ translation="MNRFTIADASKICIGRTCEVACVVSQENODCASLTPETELPRI
HVIKGVASTVTVRCCEADPCANVCNRAISRKGVKTHVWQECICQKTCVAVCPVG
AMEVYVPEVINSAGAGLVNRADKAEANKCDLCNREBDGPACMAACPTHALICVDRLKL
EOLSAEKRRRTALMF"
misc_feature 10433..14724
/ note="similar to GenBank Accession Number M73326
(ECOASCPFG)"
CDS       complement(10558..11571)

```

Alignment Scores:

Pred. No.:	Score:	Length:
Percent Similarity:	213.50	72221
Best Local Similarity:	12.07%	55
Query Match:	25.74%	137
DB:	12.74%	60
	1	12

US-10-018-706-2 (1-322) x ECU29579 (1-72221)

Oy 21 llepneagiValIleThrCysIleLeuAlaGlyCysAlaSerIleProThyTrasn 40
 DB 40873 GTCTACAGCGAGCAACCGAATTCACCGGTGACAGCAACCAATTCAGGCTACTCA 40814

Oy 41 serthserglyserglyserhiiaerthserglyserglyglyleu-Alaileglyse 60
 Db 40813 ACAACCGCAATTCACGCCAGTGCACGCCAGTACACGCCGTATACATGAAAAACGG 40758
 Oy 60 rginValilethrAspserGln-----GlyValPro-----AsnAr 72
 Db 40753 ACGCATCGTGTATATACCGGTACAGTATGGAGACATTCGAAAGGACGTTATACGGCGCATAC 40694
 Oy 72 gTyrglnVallysglnGlyAspThrValSerlysllealaglnAsTyrglyLeuAsnTr 92
 Db 40693 CTATACCGGTAAAAAAGGCGACACACTTTCTATATCCGCTGATTAATCGGCACATCATTT 40634
 Oy 92 pArgGlnileGlyHisIleAsnAsneuAsnserSerTyThrIleTyThrGlyGlnTr 112
 Db 40633 CCGGACCTTGCTCCACGCGCAACATATTCACGACCAATACGCGCTAACGTTGTAG-- 40576
 Oy 112 pLeuThrLeuTrpSerGlyAsp-LeuLyValAlArgGlnuArgserIleSerSerGlyValA 132
 Db 40575 ----ACCTTCACAGTGGGTATATGCTTCGGGTACGCCAA-----T 40541
 Oy 132 snThrAlaHisrthProSerProValAlAlValGlnSerSerArpProProValGlnGln 152
 Db 40540 CACTGCGGAAATGCCATTACCC-----AGCGCGACGACGACAGACA 40499
 Oy 152 isProAlaValGlnuAsPro-----ThrProProValValVallyblybProThr- 169
 Db 40498 AGGA-GTTGTGATCAACCGCTGCACAAATATTCACCGTTGCTGTGGGTGCACACGACA 40440
 Oy 170 -----P 170
 Db 40439 TTACGTATTCGAGTCTTCGGGTACACAGAGTGTACAAATATGTCGCGACACACACAGC 40388
 Oy 170 rOthrProProValValGlnGlnProAlaProAlaProValAlThrGlnAlaProP 190
 Db 40379 CAACTGGACACACGCGTACAGCGCTGTAAACGATACGACCAACAGACAGACACGACCGCA 40320
 Oy 190 heaIaThrGly-----SerSerGlyValMetGlnPheAlaGlyTyProValGlyAlaT 207
 Db 40319 CTGTACCAAGTACATCAACCAAGTACGCGCTATCTTCACCTGCGCTGGCCG-----ACTG 40266
 Oy 207 hTrAsnProValValArgArgPheGlyThyrAlaThrValAlaGlySerThrValThrSerA 227
 Db 40265 AGGGCAAAAGTATGAAACCTTTGGC--GCTTGTAGGGGGGC-----AACCA 40221
 Oy 227 snGlyMetTrpPheSerGlyArGAspGlyAspLeuIleAsnAlaSerAsnAlaGlyThyV 247
 Db 40220 AGGGGATTTGATATGCGACGACGACGAAAGACAGCAAGATTAATCGGACCGCAGATGGCGGG 40166
 Oy 247 alIleGlnAlaAspHisAsnMetAspGlyAlaSer-----IleValIleGlnHisThrA 265
 Db 40160 TTGCTTTATGTGTGTAAACCGCTGCGCGGTACCGGTATTCGATTATCATCAATCAAACTAA 40101
 Oy 265 snGlyPheValaSerSerTyThrIleHisIleAspAlaGlnValLyTyThrGlyAspThrV 285
 Db 40100 ATGATTTACCTGAGGCTTACGCCCATACACACACATATGCTGTGTCCGGGAAACAAGAAG 40041
 Oy 285 alArGThrGlyGlnArgIleAlaSerMetLyAsnGlnProSerGlyAlaAlaLeuPheG 305
 Db 40040 TTAAGGGCGGGCAAAAATAGCAGCACTGGGTAGACACCGGACACAGTTCAACACGCTTGC 39981
 Oy 305 luPheArgIleSerArGAsnGlyValTyValAspProLeuThrValLeu 321
 Db 39980 ATTTTGAATTCGTTACAAAGGGCAAAATCCGTAACCCGCTGCTATTTTG 39931

RESULT 32
 AP002562/c 270365 bp DNA linear BCT 07-MAR-2001
 LOCUS Escherichia coli O157:H7 DNA, complete genome, section 13/20.
 ACCESSION AP002562 GI:13362858
 VERSION AP002562.1 GI:13362858
 KEYWORDS Escherichia coli O157:H7 (strain:O157:H7, sub_strain:RIMD 0509952)
 DNA.

ORGANISM	Escherichia coli O157:H7 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.
REFERENCE	1 (sites)
AUTHORS	Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S., Kurokawa,K., Ishii,K., Hattori,M., Tatsuno,I., Abe,H., Iida,T., Yamamoto,K., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and Shinagawa,H. Complete nucleotide sequence of the prophage VT2-Sakai carrying the verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7 derived from the Sakai outbreak Genes Genet. Syst. 74 (5), 227-239 (1999)
JOURNAL	20198780
MEDLINE	2 (sites)
REFERENCE	Ohnishi,M., Murata,T., Nakayama,K., Kuhara,S., Hattori,M., Kurokawa,K., Yasunaga,T., Yokoyama,K., Makino,K., Shinagawa,H. and Hayashi,T. Comparative analysis of the whole set of RNA operons between an enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an Escherichia coli K-12 strain MG1655 Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
TITLE	3 (sites)
JOURNAL	20557356
MEDLINE	3 (sites)
REFERENCE	Yokoyama,K., Makino,K., Kubota,Y., Matnabe,M., Kimura,S., Yutsudo,C.H., Kurokawa,K., Ishii,K., Hattori,M., Abe,H., Iida,T., Yamamoto,K., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and Shinagawa,H. Complete nucleotide sequence of the prophage VT1-Sakai carrying the Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli O157:H7 strain derived from the Sakai outbreak Gene 258 (1-2), 127-139 (2000)
AUTHORS	4 (sites)
JOURNAL	20564182
MEDLINE	4 (sites)
REFERENCE	Hayaishi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K., Yokoyama,K., Han,C.-G., Ohtsubo,B., Nakayama,K., Murata,T., Tanaka,M., Toke,T., Iida,T., Takami,H., Honda,T., Sasakiwa,C., Ogasawara,N., Yasunaga,T., Kumata,S., Shiba,T., Hattori,M. and Shinagawa,H. Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic comparison with a laboratory strain K-12 DNA Res. 8 (1), 11-22 (2001)
TITLE	5 (bases 1 to 270365)
JOURNAL	21156231
MEDLINE	5 (bases 1 to 270365)
REFERENCE	Ohnishi,M., Kurokawa,K., Makino,K., Yasunaga,T., Shinagawa,H. and Hayaishi,T. Direct Submission Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871, Japan (E-mail:kengen-info.osaka-u.ac.jp, URL:http://www.gen-info.osaka-u.ac.jp/, Tel:81-6-6879-8365, Fax:81-6-6879-2047)
AUTHORS	location/Qualifiers
TITLE	1..270365
JOURNAL	/organism="Escherichia coli O157:H7"
MEDLINE	/strain="O157:H7"
REFERENCE	/sub_strain="RIMD_050952"
AUTHORS	/db_xref="taxon:83334"
TITLE	complement(435..1220)
JOURNAL	/gene="EC63388"
MEDLINE	complement(435..1220)
REFERENCE	/gene="EC63388"
AUTHORS	/note="similar to SEEB_ECOLI_g11788871 percent identity 98 in 261 aa (Conserved in E.coli K-12)"
TITLE	/codon_start=1
JOURNAL	/evidence=not_experimental
MEDLINE	/transl_table=11
REFERENCE	/product="enhanced serine sensitivity"
AUTHORS	/protein_id="BA836811.1"
TITLE	/db_xref="GI:13362859"
JOURNAL	/translation="MTVMSEKTKVLEIDLEKAATPRNRPAPFTLIESVTVAGTAA QGEAVDSALDLQHMEKEDEGTSTVPFTSLQAQQVVEEQARFVNPFVTLEMTGL ETFLVANKPLPGKEFMPEKTSILIGEGGNPISSEITLEGESILTSVAPPAMQDS


```
/note="Pfam match to entry PF00588 Spou_methylase, Spou
RNA methylase family, score 181.50, E-value 1.4e-50"
RBS
2244..2248
/note="possible RBS"
gene
2254..2673
/gene="STY2842"
/note="trix; yfig"
2254..2673
/gene="STY2842"
/note="STY2842"
2254..2673
/note="Orthologue of E. coli yfig (THI2_ECOLI); Psta hit
to THI2_ECOLI (139 aa), 94% identity in 139 aa overlap"
CDS
/codon_start=1
/product="thioredoxin 2"
/protein_id="CAD05833.1"
/db_xref="GI:16503808"
/db_xref="SPTREMBL:Q8Z4U6"
/translation="MTVCTHCOAINRIPGRLQDAACGRCGHELFDEGINATGET
LDKLKDLPLVDFWAPWCGPCNPAIFEDVBERGKVRFKVNTAEARELSARF
GIRSIPIIMIRKHGOVDMINGAVKAPFDSMLNEAL"
2353..2670
/gene="STY2842"
/note="Pfam match to entry PF00085 thired, Thioredoxin,
score 116.00, E-value 1.1e-33"
misc_feature
2419..2475
/gene="STY2842"
/note="PS00194 Thioredoxin family active site"
2746..3426
/gene="STY2843"
2746..3426
/gene="STY2843"
/note="Orthologue of E. coli yrip_ECOLI; Psta hit to
yrip_ECOLI (232 aa), 88% identity in 226 aa overlap"
CDS
/codon_start=1
/product="conserved hypothetical protein"
/protein_id="CAD05834.1"
/db_xref="GI:16503809"
/db_xref="SPTREMBL:Q8Z4J5"
/translation="MTNNAVLQRAERLARATPELARGNVRRCORLLPIKSLCLD
LTPEQAKSRCLVMEDETPMKPSNTGRLADIDTAYAFQMSRTPEQALLETVOH
DYQPIVPPASVAGEARVETSTPPAGKPLFLMDGTWPEAKMRKSPYDHDIEVIS
VDLSRLSAYRIEIRHABQYCTAEVAIALDLADGPEASISGEHPIFKRYIAGKT
QHPGNVTA"
3480..6140
/gene="STY2844"
3480..6140
/gene="STY2844"
/note="similar to proteins from bacteria, archaea and
eukaryotes e.g. Entamoeba histolytica acetyl-CoA
synthetase TR:Q9NAT4 (EMBL:AF286346) (713 aa) Psta
scores: E(): 0.34,1% id in 721 aa and Giardia lamblia
acetyl-CoA synthetase TR:Q9YIN2 (EMBL:AF107206) (726 aa)
Psta scores: E(): 0.30,0% id in 700 aa
Orthologue of E. coli yfig_ECOLI; Psta hit to yfig_ECOLI
(886 aa), 92% identity in 886 aa overlap"
CDS
/codon_start=1
/translation="putative acyl-CoA synthetase"
/protein_id="CAD05835.1"
/db_xref="GI:16503810"
/db_xref="SPTREMBL:Q8Z4J4"
/translation="MSQGLKALRPKSGIAVYAGSKPRRATYLMRNILAGFNGPV
LPTPAWKAVLVGMAPDIALSPFPDILITCNASRNIALDALGKGCCTITISA
PTSOHELLASARHYRMRLIGFNSJGLAPWQGLNARSFVPIKQKLAFTISQAVS
NITLWAOQREWFYSYFIALGSDSLIDVDELIDYARDESKTALILYELQSDARFV
SAARSAKRPKLVTKSGSPAAQRLNLSAGMDPAWDIAIPRAGILRVODTHEPSA
VETLSHMRPLHGDRLMITSNGAPALALDELMSRNGKATLSSETCLOLRTLLPHI
DIANPLDCLDDDSSEHYVKTLLILASODFALMVHSASAPGTSSEHALLETIKR
HRRGKFTVLTWCGEFSSQEARLSEGLPTPRPEGTITAFPMHVEYRNQQLR
EPALPSNLTNTAEANHLQRIAGASLDTHEQPLTHAFPMHVEYRNQQLR
AHTAEQGPVALKLRSPDIPKSEVQCVMLYLRTPASVQQAANAIPFRVKAQPOA
RTHGLIVQSMARAGAEIRLVVEHDPVFGPLIMIGEGVEWRPEBOAVVALPPLMN
```

```
LARYLVQIKQKIRARSALRPDLIVGLSQLIVQVSNLIVDCEPIORLDIHPLASA
SEPTALDVLTDIAPPDGNSRLAVPYPHOLEVEMKNGRCLEPILIPDEPOLR
QFYAOTKEDLYRFPSEINEFHEDLANTQIDYREMAFVARRMDNAEELIGVTR
AISDPNVDAEPVALVRSDLKGLGRIMERKLIAYTRDHGLKRLNGITWPNRGMVA
LARKLGPOVDIQIDEGIVGLTNLACDS"
5844..6071
/gene="STY2844"
/note="Pfam match to entry PF00583 Acetyltransf,
Acetyltransferase (GNAT) family, score 59.00, E-value
2.6e-14"
6255..7610
/gene="STY2845"
6255..7610
/gene="STY2845"
6255..7610
/gene="STY2845"
/EC_number="2.7.8.8"
/note="Orthologue of E. coli psaA (PSA_ECOLI); Psta hit
to PSA_ECOLI (451 aa), 94% identity in 451 aa overlap"
CDS
/codon_start=1
/product="CDP-diacylglycerol-serine
O-phosphatidyltransferase"
/protein_id="CAD05836.1"
/db_xref="GI:16503811"
/db_xref="SPTREMBL:Q8XFW5"
/translation="MLSKFPRKHQOHLAQLPKISQSVDDVEFYTPATRETLLEKI
ASATQRICIYALYEDPDGKGLIDALYAAKRPBELDVAVLDMHMAQGRIGAAAS
NTWADWCRLAENPGIDVPYVPIINTREALGVHFKGFIIDDSVLVSGASINDVYL
HOHDKRYDRYQOLIRNRQADIMFDWOTONLNGRGVNRFLDNTORPEKSPKINDIRLY
ROELRDSYHFOGQANDBSQVPLVGLGSSSLNKTIFHLMCAEHLKTICTPEFL
PAVLVNIIOLEBDGKVEEIIIGDKTNDYIEPDEPKFIIGALPYVEINLRFELR
LOYVNTDQVVRVLRMDQDNTYHLKGMWVDKXNMLTGNMILNPRAMLDLENLIIHD
PKQELARQREKRELEIRHTHTIYVGHYDLSIADVPKIVAKLIRLRIRIRIDRLISRI
I"
6651..6731
/gene="STY2845"
misc_feature
Alignment Scores:
Pred. No.: 0.139 Length: 274050
Score: 210.00 Matches: 84
Percent Similarity: 38.67% Conservative: 56
Best Local Similarity: 23.20% Mismatches: 138
Query Match: 12.57% Gaps: 84
DB: 13
US-10-018-706-2 (1-322) x AL627276 (1-274050)
QY 30 LeuAlaGlyCyAlaSerIys-----ProThrIYAsnSerThSergIy 44
Db 215187 CTGGCAGGTTGTACCAAGTTCGTCAAAACCCGCCGCCGTCACGTCTGTGACAGCG 215128
QY 45 SerGlySerHisArgThSergIy-----SergIyGlyLeu 56
Db 215127 TCGTCATCAATATCAATTCGGGATGTGTATCAACGCCGCCCAAAATGGCGCGAG 215068
QY 57 -----AlaIleGlySer 60
Db 215067 CCGCAACAAGCCGCCCAATTCACACCGTTCAGCGTCTGTACGACGACCCATGCAACT 215008
QY 61 GlnValIleThrAspSergIln----- 67
Db 215007 CAACCGGACGAGCAACAACCCGTGACATGGAATAGGCGCAATCGTATACATCGCCAG 214948
QY 68 -----GlyValPro-----AsnArgTYrGlnValIyGlnGly 78
Db 214947 TATGGCAATATTCGAAAGGTAGTATACGGGCGGCACTTACACCGTGAATAAGG 214888
QY 79 AspThrValSerIysIleAlaGlnArgTYrGlyLeuAsnTPARGIuIleGlyHisIle 98
Db 214887 GATACGCTTTTAAACGCTGATTCACGGGAATATTCGGGATCTGGCCAGAGA 214828
QY 99 AsnAsnLeuAsnSerSergIYThrIleTYrThrGlyGlnTPleu----- 113
```

Db 214827 AACGATATTCCTGCCCCGTATAGCTGATAGGCAAACTTTGCAAGTTGTACGCA 214768
 Qy 114 -----ThleuTpsrGlyAspLeuValArgGluArgSerIleSerSergly 130
 Db 214767 TCCGGTACCGCAATTAATCTGCGGAAATGCGATCACCAGCGGATGACGACAGCAAGGA 214708
 Qy 131 ValAsnThrAlaHisThrProSerProValAlaValGlnSerSerArgProProVal--- 149
 Db 214707 GTTGACACGAGTCTGCAAAATTCACCGTTGACGTGCGTCTCAACCAATTAACG 214648
 Qy 150 -----GlnGlnHisProAlaValGlnLysProThr-----ProPro 161
 Db 214647 TATTCGAGGCGCTCAGGTGCAAAAGTCTTAACAAATGTTGCAAAACAAAGCTTCT 214588
 Qy 162 ValValValValLysLysProThrProThrProProValValGlnGlnProAlaProVal 181
 Db 214587 GGGACGTTTCTACAGCGCTCTGTAACGCGACCAACGGTAAACGACCAACCAATGCA 214528
 Qy 182 AlaProProValThrGlnAlaProPheAlaThrGlySerSerglyValMetGlnPheArg 201
 Db 214527 AGCAGTACGTCAACCAAGCGCGCG-----ATTCCGACATGCGCG 214489
 Qy 202 TyrProValGlyAlaThrAsnProValValArgArgPheGlyThrAlaThrValAlaGly 221
 Db 214488 TGGCCG-----ACGATGCGCAAAAGTATGCAAACTTTGGC---GCTTCGCAAGCGCGC 214438
 Qy 222 SerThrValThrSerArgLysMetTyrPheSerglyValArgAspGlyAspLeuIleAsnAla 241
 Db 214437 -----AATAAGGGATTCGACATTCGACGCGAGTAAAGGACAGCGCTATCTGCGCA 214390
 Qy 242 SerAsnAlaGlyThrValIleGlnAlaAspHisAsnMetAspGlyAlaSer-----Ile 259
 Db 214389 ACCGCGATGGCGCGCGTATATGCGCGTAACGCACTGGTGTATACGTAATCTTAAT 214330
 Qy 260 ValIleGlnHisThrAsnGlyPheValSerSerglyIleHisIleLysAspAlaGlnVal 279
 Db 214329 ATCATCAACATCAACATGATTCCTGAGTCCATCCATCAATGATGATGATGATGATG 214270
 Qy 280 LysThrGlyAspThrValArgThrGlyGlnArgIleLysSerMetLysAsnGlnProSer 299
 Db 214269 CCGGAACACAGAGAGTAAAGCGGGGCAAAATCCCTACTATGGGTACACCGCGACC 214210
 Qy 300 GlyAlaAlaLeuPheGlnPheArgIleSerArgAsnGlyValTyrValAspProLeuThr 319
 Db 214209 AGCTCAACACGCTTGAATTGTAATTCGTACAGGGAATCCGTAACCCGCTCGT 214150
 Qy 320 ValLeu 321
 Db 214149 TATTTA 214144
 RESULT 34
 SDU6131
 LOCUS SDU6131 1290 bp DNA linear BCT 24-FEB-1999
 DEFINITION Salmonella dublin nlpd gene, strain lane.
 ACCESSION AJ006131
 VERSION AJ006131.1 GI:3150245
 KEYWORDS nlpd gene.
 SOURCE Salmonella enterica subsp. enterica serovar Dublin.
 ORGANISM Bacteri; Proteobacteria; gamma subdivision; Enterobacteriaceae; Salmonella.
 1 (bases 1 to 1290)
 Paesold, G.
 TITLE Direct Submission
 JOURNAL Submitted (18-MAY-1998) Paesold G., Institute of Microbiology,
 Swiss Federal Institute of Technology, Schmelzbergstrasse 7, ZH,
 8092, SWITZERLAND
 2 (bases 1 to 1290)
 Paesold, G. and Krause, M.
 TITLE Analysis of rpos mRNA in Salmonella dublin: identification of
 multiple transcripts with growth-phase-dependent variation in
 transcript stability
 JOURNAL J. Bacteriol. 181 (4), 1264-1268 (1999)

MEDLINE 99138748
 PUBMED 9973354
 FEATURES
 source
 Location/Qualifiers
 1..1290
 /organism="Salmonella enterica subsp. enterica serovar
 Dublin"
 /strain="lane"
 /db_xref="taxon:98360"
 120..1253
 /gene="nlpd"
 120..1253
 /gene="nlpd"
 /codon_start=1
 /protein_id="CA006881.1"
 /db_xref="GI:3150246"
 /translation="MSAGSPKFTVSRIALSLVSLMAGCTSSNPAPVTSVDSGS
 SNTSGMLITTPPKGATTOQPOAPOLQPVORPVTPQVOTPVTEOPVOMENGRIV
 YNRQYGNIPKSGSYTGSTVTKKGDILFYIAMITGDFDLAQRNLSISAPVSLNGOT
 LQVNASGMPITGNAITQADANQGVTRSAQNSVAVASOPTIYISBEGSGANAK
 MLPNNKPAQTVVTAAPVTAAPVSTBENASSTCSAPISANRMPETDKVINEPQASBG
 NKSIDYAGSKQALVATADGRVYAGNALRGYNLIIRKNDVLSYAHNDVTLVRE
 QDEVKAGQKATATGSGTSTRLHFIRKKGKSVNPLRYLPOR"
 BASE COUNT 344 a 326 c 335 g 285 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 0.000432 Length: 1290
 Score: 209.00 Matches: 84
 Percent Similarity: 38.25% Conservative: 56
 Best Local Similarity: 22.95% Mismatches: 138
 Query Match: 12.51% Indels: 88
 DB: 1 Gaps: 13
 US-10-018-706-2 (1-322) x SDU6131 (1-1290)
 Qy 30 LeuAlaGlyCysAlaSerLys-----ProThrTyraSerThrSergly 44
 Db 186 CTGGCAGGTTGTACCACTGCTGTAACACCTCCGCGCGCTGACGTTGTGACAGTGGG 245
 Qy 45 SerGlySerHisArgThrSergly-----SerglyGlyLeu 56
 Db 246 TCGTCATCAAAATCAACATTCGGGATGTTGATCACACCGCGCAAAATAGGCGGACG 305
 Qy 57 Ala----- 57
 Db 306 ACGCAACAAACCGCGCAACAGCGCGCAATTAACACCGCTTACGCTCTGTAACGACG 365
 Qy 58 ---IleGlySergLysValIleThrAspSergL----- 67
 Db 366 CCGATGCAAACTAACCGGTACCGGACCAACCCGTCGATGAGAAATGCGCGAATTGTA 425
 Qy 68 -----GlyValPro-----AsnArgTyrGln 74
 Db 426 TACATCGCCAGTATGGAATATTCGAAAGTATGAGTATACGGGTGGCAGTACTTACACC 485
 Qy 75 ValLysGlnGlyAspThrValSerLysIleAlaGlnArgTyrGlyLeuSntPArgL 94
 Db 486 GTGAAAAAGGCTATACGCTTTTTCATCGCTGATTAACCGGAAACGATTTCCCGGAT 545
 Qy 95 IleGlyHisIleAsnLeuAsnSerSergTyrThrIleTyrThrGlyGlnTyrLeu--- 113
 Db 546 CTGGCCCAAGAAACAGTATTTCTGCGCGGTATAGCTGAATTAAGGCAAACTTGGCAA 605
 Qy 114 -----ThleuTpsrGlyAspLeuValArgGluArgSer 126
 Db 606 GTTGTAACGATCCGATATGCAATTAATCTGCGGAAATGCGATCACCAGCGGATGCA 665
 Qy 127 IleSerSergLysValAsnThrAlaHisThrProSerProValAlaValGlnSerSerArg 146
 Db 666 GCACACAGAAAGAGTGTGACAGGCTGCAAAATTCACCGTTGACGTGCGTCTCA 725
 Qy 147 ProProVal-----GlnGlnHisProAlaValGlnLysProThr 159

[illegible]

OY	93	ArgGluIleGlyHisIleAsnAsnLeuAsnSerSerGlyThrIleLeuYrThrGlyGlnTrp	112
Db	1289	AAAGAAATGGCCACACATAATATATATGTCTGAGCGCAATATCATCTGAGTATGGCAACGTA	1348
OY	113	LeuThrLeuTrpSerGly-----AspLeuIysValArgGluArgSerIleSerSer	129
Db	1349	TTGAAATTCGCAAAATATATATTCGCCATAGCAATATATGATACCAACACAGACAAATTAATGAA	1408
OY	130	GlyValAsnThrAlaHisIsthTrpSerSerProValAlaValInsSerArgProProVal	149
Db	1409	TCAGAGGTGACCAAAAATACAGTCATAGAG---ACATGGAATGCTATATTAACCAACAAT	1465
OY	150	GlnGlnHisProAlaValGlnIlyProCthTrpProProValValValIlyIlySylProThr	169
Db	1466	GAACCAAAATGAACCCGTTGCTACCAACACATTCACAAATCCAAATCAATTA---1519	1519
OY	170	ProThrProProValValGlnGlnProAlaProValAlaProProValTrhGluAlaPro	189
Db	1520	---ACACCTCCAGCCACCTCAAAATATGCTTGATTTGGCCA-----1558	1558
OY	190	PheAlaThrGlySerSerGlyValMetGlnPheArgTyProValGlyAlaThrAsnPro	209
Db	1559	-----ACAAAATGAGAAA1570	1570
OY	210	ValValAlaArgArgPheGlyYrThrAlaThrValAlaGlySerThrValThrSerAsnGlyMet	229
Db	1571	ATTATTCAGAGATTTTCCAGTGCTGATGAGAGCC-----AATAAGATGATT1615	1615
OY	230	TrpPheSerGlyArgAspGlyAspLeuIleAsnAlaSerAsnAlaGlyYrThrValIleGln	249
Db	1616	GATATTACCGCTTCCTGACCAACAGCTGTATATGACGACCTGCTGACAGATGCTATAT	1675
OY	250	AlaAspHisAsnMetAspGlyAlaSer-----IleValIleGlnHisIsthAsnGlyPhe	267
Db	1676	GCCCGAGACGCTTTACCGTGAATGATGTGTAATTATATTAATTAACCAATTAATGACAGTTAT	1735
OY	268	ValSerSerGlyTrpIleHisIleIlyAspAlaGlnValIlySthrGlyAspThrValArgThr	287
Db	1736	TTAAGTGCTTATGACCACTAATGAAGAAATGATCTGCTCAAAATCAGCAAGAAAGTAAAGCG	1795
OY	288	GlyGlnArgIleAlaSerMetLysAsnGlnProSerSerGlyAlaIleAlaLeuPheGluPheArg	307
Db	1796	GGTCAACAAATGTCTTAAATGGGAAGTTCTGGAACAACAATCAATCAATCCATTTTGAA	1855
OY	308	IleSerArgAsnGlyValIlyValAspProLeuThrValIleu321	321
Db	1856	ATTGCTTATTAAGTCAATCACTGATGATCAATGATGATTTTA1897	1897
RESULT 36			
AE008833/c			
LOCUS	20984 bp	DNA	linear BCT 31-Jul-2002
DEFINITION	Salmonella typhimurium LT2, section 137 of 220 of the complete		
ACCESSION	genome.		
VERSION	AE008833.1		
KEYWORDS	GI:16421460		
SOURCE			
ORGANISM	Salmonella typhimurium LT2.		
	Salmonella typhimurium LT2		
	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;		
	Salmonella.		
REFERENCE	1 (bases 1 to 20984)		
AUTHORS	McClelland,M., Sanderson,K.E., Spieth,D., Clifton,S.W.,		
	Latrelle,P., Courtney,L., Porwollik,S., Ali,J., Dante,M., Du,F.,		
	Hou,S., Layman,D., Leonard,S., Nguyen,C., Scott,K., Holmes,A.,		
	Grawal,N., Mulvaney,E., Ryan,E., Sun,H., Flores,L., Miller,W.,		
	Stoneling,T., Nhan,M., Waterson,R. and Wilson,R.K.		
TITLE	Complete genome sequence of Salmonella enterica serovar Typhimurium		
JOURNAL	Nature 413 (6858), 852-856 (2001)		
MEDLINE	21534948		
PubMed	11677609		
REFERENCE	2 (bases 1 to 20984)		

AUTHORS The Salmonella typhimurium Genome Sequencing Project.
TITLE Direct Submission
JOURNAL Submitted (29-MAR-2001) Genome Sequencing Center, Department of
Genetics, Washington University School of Medicine, 4444 Forest
Park Boulevard, St. Louis, MO 63108, USA
COMMENT Supported by NIH grant 5U 01 AI43283

Coding sequences below are predicted from manually evaluated
computer analysis, using similarity information and the programs;
GLIMMER: <http://www.tigr.org/softlab/glimmer/glimmer.html> and
GeneMark; <http://opal.biology.gatech.edu/GeneMark/>

EC numbers were kindly provided by Junko Yabuzaki and the Kyoto
Encyclopedia of Genes and Genomes; <http://www.genome.ad.jp/kegg/>,
and Pedro Romero and Peter Karp at EcoCyc;
<http://ecocyc.PangeaSystems.com/ecocyc/>

The analyses of ribosome binding sites and promoter binding sites
were kindly provided by Heladia Salgado, Julio Collado-Vides and
ReguonDB;
http://kinich.cifn.unam.mx:8850/db/reguondb_intro.frameset

This sequence was finished as follows unless otherwise noted: all
regions were double stranded, sequenced with an alternate
chemistries or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one ml3 subclone.

FEATURES
source
1..20984
/organism="Salmonella typhimurium LT2"
/strain="LT2; SGSC 1412; ATCC 700720"
/db_xref="ATCC:700720"
/db_xref="taxon:99287"
/note="LT2"
71..1003
/gene="STM2912"
71..1003
/note="STM2912"
/note="similar to E. coli putative transcriptional
regulator lysR-type (AAC7313.1); Blastp hit to AAC7313.1
(304 aa), 32% identity in aa 6 - 289"
/codon_start=1
/transl_table=1
/product="putative lysR family transcriptional regulators"
/protein_id="AAL21792.1"
/db_xref="GI:16421461"
/translation="MTVKEIAMLNORMSLFIAVDSGSPFAAASGQTRAVSFNI
ROLEKELGVTLLSTRRLTLTDAGVLFYQGVNLLNAKLNODEVRASHSGGELR
ITTPPEGEVITPLAOFOSRHPRLRIHSSSHHADLIERPVALRIGLSADRY
RALISRPTIIPVAPQVLAHPVPSLSLESLAQBHTHERLPTPRATVTSNCHSR
LEISKAGKISVDSASLMAFALAGSGVALLPQWLVNLALEDGTLIHVLPDHPROGI
YAVYPDAHVSTKRAFIIDFLRSQDCGSHAPS"
complement (1045..2548)
/gene="STM2913"
complement (1045..2541)
/gene="STM2913"
/note="similar to E. coli putative transport protein
(AAC75782.1); Blastp hit to AAC75782.1 (454 aa), 33%
identity in aa 11 - 446"
/codon_start=1
/transl_table=1
/product="putative permease"
/protein_id="AAL21793.1"
/db_xref="GI:16421462"
/translation="MTSLPTPIIGLIYAVFLVWLVTNRVHALIAMLAAACIAGLLG
GMGIDKLTSLTSGGTGSLIGLIVIGVMMGRSLLEVSAAERLAFIMVLTHTTP
EMALATGYIVSIPFVDSAFVILPVKAKALAGSRLTLGVALAGGLVNTGRRV
PTPGIAGVIGFVNDIGAMLLTGMAALPCVIGIYFAWMLDKYRDPVPTIADRY
NMALEQYKEKEKELPSLTSLPVIPIVILFIKALICSTLATEGSGIATPPEVO
AINTVGSFVIALAISVLAVYTLVPRMDKHTAERLEGLQSAIGILLVAGAGLGA
ILRDSGAQQLAEQVANIIPISPIIPVATLVRLIQSGTVMAMITASISAPILAQI
PGIMMLAAQAATGSLFFGYFNDLSLFWVAVRMKGVSDVKQWVMSVPTTIAAIGG

gene
complement (2538..3495)
/gene="STM2914"
complement (2538..3488)
/gene="STM2914"
/note="similar to E. coli UDP-galactose-4-epimerase
(AAC73846.1); Blastp hit to AAC73846.1 (338 aa), 25%
identity in aa 1 - 162"
/codon_start=1
/transl_table=1
/product="putative nucleoside-diphosphate-sugar epimerase"
/protein_id="AAL21794.1"
/db_xref="GI:16421463"
/translation="MQIITGGGFLGQKLASALLNSLARNELLVCLKMPARLSDS
PRACLEADLPQGVESVITRANSTVYTHLAATYSNADDPDLGKAKNLDITRRLD
ACRQPOKIRFVSSSLAVGTTIPECTITATLPRSYAQAKACLVNDYTRKG
YVDGLARLPPTICVRGPKRNPASVSAIIRPLQGTTCVPSSESLMISATV
IHNLSLATLPAPGEASINLPGISVWGEMLETLCOAGCOARDRVTHORDEGEKI
VASWPRIDNORALALGFADRFDDIIRFRODMEGRS"
complement (2543..2548)
/gene="STM2914"
complement (2543..2548)
/note="putative RBS for STM2913; RegulonDB:STMS1H002970"
complement (3490..3495)
/gene="STM2914"
/note="putative RBS for STM2914; RegulonDB:STMS1H002971"
complement (3528..4318)
/gene="ygbM"
/note="STM2915"
complement (3528..4304)
/gene="ygbM"
/note="similar to E. coli orf, hypothetical protein
(AAC75781.1); Blastp hit to AAC75781.1 (258 aa), 79%
identity in aa 1 - 258"
/codon_start=1
/transl_table=1
/product="putative endonuclease"
/protein_id="AAL21795.1"
/db_xref="GI:16421464"
/translation="MPRRANLSMFTPEAPTEPFAAABEGPQVETLPPYDRAE
IKALSHNLTIALFNTSAGDTAAGWGRALPERHDARDIDLAEYALALCEOV
HIMGVVDPDGDARVATFIDNIRVADARADKILBALSPGVKGLFSSQYO
ALGIAEVDYRNVLIQDTEFAQVVDNLSLILEYVGRVYHVOIASLPDRHEDDGE
INYPMLPRLFDVDVGRYCGEYQPRNTYDQGLGFWFAMR"
complement (4309..4962)
/gene="ygbL"
/note="STM2916"
complement (4309..4947)
/gene="ygbL"
/note="similar to E. coli putative epimerase/aldolase
(AAC75780.1); Blastp hit to AAC75780.1 (212 aa), 85%
identity in aa 1 - 211"
/codon_start=1
/transl_table=1
/product="putative fucose phosphate aldolase"
/protein_id="AAL21796.1"
/db_xref="GI:16421465"
/translation="WTLAKBEHALREBMRIASFQRGVATSGAGNLSLLPDGNT
LATPTGSLGNDLPQRLSKVDPOGEMLSGKPSKVRPHLALYNNPCCKAVHLST
WSTALSGCLGNDPQNVIRPFTPYVMMGIPLVYRPGDRRIARDLAALAAHQA
PLAHGVVCGENIQEAAANTELEETKALIFIGERIRYLTTETIAQNR"
complement (4313..4318)
/gene="ygbL"
/note="putative RBS for ygbM; RegulonDB:STMS1H002972"
complement (4944..6217)
/gene="ygbK"
/note="STM2917"
complement (4944..6206)
/gene="ygbK"
/note="similar to E. coli orf, hypothetical protein
(AAC75779.1); Blastp hit to AAC75779.1 (388 aa), 79%
identity in aa 1 - 384"
/codon_start=1
/transl_table=1

```

/product="putative RNA synthase"
/protein_id="AA121797.1"
/db_xref="GI:16421466"
/translation="MLKIGVADDFGTGATDIASFLVENGMPYVOINDVPTGTPEGCD
AVVSLKTRSPGPAEAIKOSLALVWLKQCOOVFPYKSTPSTAEKNGIPYTDAL
MVALDTSFTVPALPVNGRTVYGYLFVNMHLAESGMHRHPINPTDYLPRIMEA
OAGRCGVTPAOTIDEGVAATRALASLQDEGYRAYVLDLNERHLEIOGVLEADPL
VGGSGGAMGLROMAGVSGASAGYPIPSGSAVVISGSCSMNTNOOAFYRQHAP
RDVAVACISLSEPEAYAEALAOVLSDPSLAPMISATSLTOLAIAIOOYGTAS
HAYVALSLAARLARAGITPPIVAGCEPVSIVOSLITGFHLGPICISREVPVWNL
HAYVALSKSNRDESFTRAGREPOV"
/complement(4957..4962)
/gene="ygbL"
/notice="putative RBS for ygbL, RegulonDB:STM1H002973"
/complement(6200..7134)
/gene="STM2918"
/notice="ygbJ"
/complement(6200..7123)
/gene="ygbJ"
/notice="similar to E. coli putative dehydrogenase
(AAC75778.1); Blastp hit to AAC75778.1 (302 aa), 82%
identity in aa 1 - 299"
/codon_start=1
/transl_table=11
/product="3-hydroxyisobutyrate dehydrogenase"
/protein_id="AA121798.1"
/db_xref="GI:16421467"
/translation="MTGTDFHVGIVLGSMMGMGAARSCLRAGLSTMGADINPOCAN
LLRAGCGAASAREPAGVVDVAVLVVNAAOVRVFGEDGVAHMLKPGASVWVST
ISSADQEIATLALINLNMLDAPVSGAVVAAGSEMTVMASGSAAPATRLKPVAV
ASNYRISLPTGAGSTYKTIHILAGVHIAAAREMALAARAGIPIDVWYVYTHAG
NSMPEFMQHVVDYTPRRAVDLFFVMDLGVADTARALRPLPLASTALNMTSAS

```

Alignment Scores:

```

Pred. No.: 0.00938 Length: 20984
Score: 209.00 Matches: 84
Percent Similarity: 38.25% Conservative: 56
Best Local Similarity: 22.95% Mismatches: 138
Query Match: 12.51% Indels: 88
DB: 1 Gaps: 13

```

US-10-018-706-2 (1-322) x AE008833 (1-20984)

```

QY 30 LeuAlaGlyCyAlaSerIys-----ProThrIysAsnSerThrSergIy 44
DB 13112 CTGGCAGGTTGTACCACTCTCCGCGCGCGTCACTGTGTGACAGTGGT 13053
QY 45 SerGlySerHisArgThrSergIy-----SergIyGlyLeu 56
DB 13052 TCGTCATCAATATACCAATCCGGGAAATGTATACACACCGCGCAAAATGGCGCGACG 12993
QY 57 Ala----- 57
DB 12992 ACGCAACAACCGCGCAACACCGCGCAATTCACCCGTTACGGTCTGTAAACGACG 12933
QY 58 ---IleGlySerGlnValIleThrAspSergIn----- 67
DB 12932 CCCATCAAACTCAACCGGCGGCAACCAACCGGTCAATGAGAAATGGCGCAATCGTA 12873
QY 68 -----GlyValPro-----AsnArgIyGln 74
DB 12872 TACAAATGCCAGTATGGGAATATTCGGAAGAGTAACTATACGGGTGGCACTTACACC 12813
QY 75 ValIyGlnGlyAspThrValSeriIysIleAlaGlnArgIyGlyLeuAsnIleArgGln 94
DB 12812 GTGAAAAGAGGTATACGCTTTTTCATCGCTGTGATTAACCGGGAACGATTTCCCGCAT 12753
QY 95 IleGlyHisIleAsnIleuAsnSerSeriIyThrIleIyThrGlyGlnIlePleu--- 113
DB 12752 CTGGCCCAAGAAAGATATTTCTGCCCGGTATAGCTGAATGAGGCAAACTTTGCAA 12693
QY 114 -----ThrLeuIyPserGlyAspIleuIyValArgGlyIyGser 126

```

```

DB 12692 GTTGTAAACGCAATCCGATACGCCAATTACTGCGCGGAATGCCATCACCCACGGCGATGCA 12633
QY 127 IleSerSergIyValAsnThrAlaHisIleThrPserSeriProValAlaValGlnSeriArg 146
DB 12632 GCACAGCAAGAGATGTGGACAGATGTGCAAAATATCCACCGTTGCAAGTTGGCTTCAA 12573
QY 147 ProProVal-----GlnGlnHisProAlaValGlnIySProThr 159
DB 12572 CCAACATTCATCGATTCTGAGGCTCAGGTGAACAAAGTCTTAACAAATATGTTGCCAAMC 12513
QY 160 -----ProProValIyValIyValIySProThrPProThrProProValIyGlnGln 177
DB 12512 AACAGCTGTGGACGTTGTTCACAGCCCTGTAAAGCAGCAACGAGTAACAGACGACC 12453
QY 178 ProAlaProValAlaProProValIyThrGlnAlaProPhelAlaThrIySeriSergIyVal 197
DB 12452 GAACCAATGCACAGATACGTACACAGCGCCCG-----ATT 12414
QY 198 MetGlnPheArgIyIyProValIyAlaIyThrAsnProValIyAlaArgPheGlyIyThrAla 217
DB 12413 TCCGATGGGCGTGGCCG-----ACGATGGCAAGATGATCCAAACTTTGGC---GCT 12363
QY 218 ThrValAlaGlySerThrValIyThrSeriSergIyMetIyPheSeriGlyArgAspGlyAsp 237
DB 12362 TCCGACAGGGGCG-----AATAAAGGATTCGACATTCGAGGCACTAGTAAGGAGCAG 12315
QY 238 LeuIleAsnAlaSerAsnAlaGlyIyThrValIleGlnAlaAspHisAsnMetAspGlyAla 257
DB 12314 GCTATCGTCCGACACCGCTGATGGCGCGCTGTATATCCGATTAACGACGCTGGTTTAC 12255
QY 258 Ser-----IleValIleGlnHisIleThrAsnGlyPheValSeriIyIleHisIleIyS 275
DB 12254 GGTAACTTATTATCATCAACAAATACATACATATTAAGTCTGAGTCCGCTTACGCCCATTAATGAT 12195
QY 276 AspAlaGlnValIySeriIyGlyAspThrValArgThrGlyGlnArgIleAlaSerMetIyS 295
DB 12194 ACGATGCTGCTGCGCGGCAACAAACAGAAATTAAGCGGGGCAAAATCGCTATATGGGT 12135
QY 296 AsnGlnProSeriGlyAlaAlaIleuPheGlnPheArgIleSeriArgAsnGlyValIyVal 315
DB 12134 AGCACGGGACCACTTACACGCTTGCATTTTAATTCGTTACAAAGGGAATCCGTA 12075
QY 316 AspProIleuThrValIleu 321
DB 12074 AACCGGCTGCGTTATTTA 12057
RESULT 37
AE004139 10104 bp DNA linear BCT 31-JUL-2000
LOCUS Vibrio cholerae chromosome I, section 47 of 251 of the complete
DEFINITION chromosome.
ACCESSION AE004139 AE003852
VERSION AE004139.1 GI:9654953
KEYWORDS
SOURCE
ORGANISM
Vibrio cholerae.
Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
1 (bases 1 to 10104)
Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwinn, M.L.,
Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Umayam, L.,
Gill, S.R., Nelson, K.E., Read, T.D., Tettelin, H., Richardson, D.,
Ernst, P., McDonald, L., Urdack, T., Fleischmann, R.D.,
Nierman, W.C. and White, O.
DNA sequence of both chromosomes of the cholera pathogen Vibrio
cholerae
JOURNAL Nature 406 (6795), 477-483 (2000)
MEDLINE 20406833
PUBMED 10952301
2 (bases 1 to 10104)
Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwinn, M.L.,
Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Umayam, L.A.,
Gill, S.R., Nelson, K.E., Read, T.D., Tettelin, H., Richardson, D.,

```

AE L D A E P D I M A L I R G N R M R H D R A I A L K A Q N L S W Q V E D Q I T L R F S L D A G S F A T S L V R

```

/gene="VC0535"
complement(6140..8728)
/gene="VC0535"

```

PID:882626 GB:U00096; identified by sequence similarity;
 putative"
 /codon_start=1
 /transl_table=11
 /product="DNA mismatch repair protein Muts"
 /protein_id="AF03703.1"
 /db_xref="GI:9654962"
 /translation="MMKSNAPSESLSHHTPMQOYLRKAENPDLIFPMGDPVEL
 FYDDAKASBELDILTRGASAGEPIPMAGVPAHVEGYLAKIMQESVYICQIG
 DATSKGPVERKIVITVPGYVDELLSEPVNLIATVYHNGPGVYMTITSGRF
 OLSEPTREEMAEIORTSPRLIPEDSPHMASSRGNNRRPIIMEFEDTQSLIDA
 NOCGTRDLVGVGPOAKGLCAAGLTIOYVOTOTKALPHIRSLTMDQDSVLIDA
 ARPNLEITHNLACGDTNTLAEDVDCATPQSRMLKWTIHQMRDNATLNGRLDIT
 EKRTNLYGEIHPVYQIDGTERILAKLRSPRPRDILRHSMAQOQLEHVSVEL
 KOPHLELTHASPMDELCDLERATIKENRPVIRGQVLAUGSNAEDERDLANGA
 TEFERLEABERDRIQDKYNNVHPYIOVSQSHLQQLAASVADLVQLNLA
 IIELEKQEDKVLNRSKALALEKQLEELFDLMPHELIQQLAASVADLVQLNLA
 EKAENLEYCRPTLVQAGIHIQGRHPVERVWNEPFIENLPORMLITTPNM
 GKSTYRQTAIALMAHIGSYVPAESASIGLDRIPTRIGASDILASGRSTFVEMT
 ETPANILHNATNSLVLMDEIGRTYDGLSLAMASAEWLAKIEMTLFATHYEELT
 ELPNVPLHANVLAVERGDIAFMHVAQEGASKSXGLAVALAGAPKPVTKNARA
 KLOQLELSSQPAETKPSRVDIANQLSLIPESAVEQALAGVDPDLPQALMLLY
 QRLKL"
 /gene=".8836
 /gene="VC0536"
 /735.".8836
 /735.".8836
 /gene="VC0536"
 /note="identified by Glimmer2; putative"
 /codon_start=1

Alignment Scores:
 Pred. No.: 0.0048 Length: 10104
 Score: 208.50 Matches: 85
 Percent Similarity: 40.308 Conservative: 48
 Best Local Similarity: 25.764 Mismatches: 198
 Query Match: 12.498 Indels: 59
 DB: 1 Gaps: 12

US-10-018-706-2 (1-322) x AB004139 (1-10104)

QY 17 ArgLeuGlyLeuIlePheGlyValIleThrThrCysIleLeuAlaGlyCysAlaSerIys 36
 DB 4049 CGACGGGGCTTATTA-----TTGTTTTCACGCTTATTTTCGGCTGATCTCA--- 4096
 QY 37 ProThrTyraSerThrSerGlySerGlySerHisArgThrSerGlySerGlyIleu 56
 DB 4097 CCTACGCCGACGACGAGTCTGTGATGGGTAAAGATTCAACAAAGTTGAGCGGTAGT 4156
 QY 57 AlaIleGlySerGlyValIleThrAspSerGlyValProAsnArgTyGlnValIys 76
 DB 4157 TACCCGCGGAATTAC-----TATCAACTTAA 4183
 QY 77 GlnGlyAspThrValSerIleAlaGlnArgTyGlyLeuAsnTrpArgGlnIleGly 96
 DB 4184 AAAGCGACGCTTATTTATTTGCTTACCTCAACAGATAAAGATGTAATGATCTGATC 4243
 QY 97 HisIleAsnAsnLeuAsnSerSerTyThrIleTyThrGlyGlnTrpLeuThrIleTrp 116
 DB 4244 AGTTCAACGATCTGCTCCACCTTACACATTCCTCTGGCAAAAATTAAGCTTGG 4303
 QY 117 -----SerGlyAspLeuIleVal----- 122
 DB 4304 TTACCTAATTATCTCTCTGCTTATGATGAGACGGCGGGGAGCTACCGTGGCTTGC 4363
 QY 123 -----ArgGlyAspSerIleSerSerGlyValAsnThrAlaHisThrProSerPro 139
 DB 4364 GCTTCACACCAAGTGCAGAGTGTCCAAAGCGCGGACGACGCGCAACGATTGCTCAACC 4423
 QY 140 ValAlaValGlnSerSerArgProProValGln-----GlnHisPro 153
 DB 4424 GTGAGCAAAAGCTTAATACCAAAAATGATACAAAATGGGTCAACGATATGCAAAATTTG 4483
 QY 154 AlaValGlnIysProThrProProValValValIysIysProThrProThrProPro 173

DB 4484 ACTAABAAAAGATCCAGTTAAGACGGTTATGCAACAAACAAAGAGATTTGGAACCT 4543
 QY 174 ValValGlnGlnProAlaProVal-----AlaProProValThrGlnAlaPro 189
 DB 4544 GTGGTAAACAAATGTTAATGATTAATGATACCAAAAGCTTAACCTTGATGAAGAG-- 4600
 QY 190 PheAlaThrGlySerSerGlyValMetGlnPheArgTyProValGlyAlaThrAsnPro 209
 DB 4601 -----ATAGCCAAAGGCGCTATGAGCCG-----ACAAAAGTGA 4633
 QY 210 ValValArgArgPheGlyThrAlaThrValAlaIleSerThrValThrSerAsnGlyMet 229
 DB 4634 GTCAATAAAGAACTTTCT-----GCGGCGCATCA--GGCAACAAAGGAGATA 4678
 QY 230 TrpPheSerGlyArgAspGlyAspLeuIleAsnAlaSerAsnAlaGlyThrValIleGln 249
 DB 4679 GACATTTGCTGGCCCAACGCTGTCAGAGCTGTCTGCTACTGACAGATGGAACCGTAGTAT 4728
 QY 250 AlaAspHisAsnMetAspGlyAlaSer-----IleValIleGlnIleThrAsnGlyPhe 267
 DB 4729 TCGGCAACGCAATTACGCGTTATGCAACCTTATCATTAATAACATATGACACTAT 4798
 QY 268 ValSerSerTyThrIleHisIleAspAlaGlnValIleThrGlyAspThrValArgThr 287
 DB 4799 TTAAGTGCCTATGCCCAATGATCAGCTGCTAGCAAGAGACAAACTGTGCAAGCA 4858
 QY 288 GlyGlnArgIleAlaSerMetIleAsnGlnProSerGlyAlaAlaLeuPheGlnPheArg 307
 DB 4859 GCCAGAAAGATCGCACATGAGAAAGTTCTGTGTTCAACAGCGTACGCTTGCACTTGA 4918
 QY 308 IleSerArgAsnGlyValTyValAspPro 317
 DB 4919 ATCCGTTATCAAGGAAAGTCAAGTCAATCT 4948
 RESULT 38
 ECORPOS
 LOCUS
 DEFINITION
 Escherichia coli gene for RNA polymerase sigma factor, ORF-X
 protein, L-Isoaspartyl protein carboxy methyltransferase, partial
 and complete cds.
 D17549
 D17549.1 GI:404097
 L-Isoaspartyl protein carboxyl methyltransferase; RNA polymerase
 sigma factor.
 Escherichia coli (strain:DH1) DNA, clone:pKTF106.
 Escherichia coli
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Escherichia.
 1 (sites)
 Tanaka, K., Takayanagi, Y., Fujita, N., Ishihama, A. and Takahashi, H.
 Heterogeneity of the principal sigma factor in Escherichia coli:
 the rpoS gene product, sigma 38, is a second principal sigma factor
 of RNA polymerase in stationary-phase Escherichia coli
 Proc. Natl. Acad. Sci. U.S.A. 90 (8), 3511-3515 (1993)
 93234527
 2 (sites)
 Ichikawa, J.K., Li, C., Fu, J. and Clarke, S.
 A gene at 59 minutes on the Escherichia coli chromosome encodes a
 lipoprotein with unusual amino acid repeat sequences
 J. Bacteriol. 176 (6), 1630-1638 (1994)
 94179096
 JOURNAL
 MEDLINE
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 INST. OF MOL. & CELL. BIOSCI.; YAYOI 1-1-1, Bunkyo-ku, Tokyo 113.

Japan (Tel:03-3812-2111 (ex.7825), Fax:03-3813-0539)

FEATURES

source

Location/Qualifiers
1..1612
/organism="Escherichia coli"
/strain="DH1"
/db_xref="taxon:562"
/clone="pKTF106"

gene

1..57
/gene="pcm"

CDS

<1..57
/gene="pcm"
/codon_start=1
/transl_table=1
/product="L-isoaspartyl protein carboxy methyltransferase"
/protein_id="BA04486.1"
/db_xref="GI:404098"
/translation="IDTVEAVFVPLVKGELA"

gene

490..1335
/gene="ORF-X"
490..1335
/gene="ORF-X"

CDS

/codon_start=1
/transl_table=1
/product="ORF-X protein"
/protein_id="BA04487.1"
/db_xref="GI:404099"
/translation="MENGRIYVNRQYGNIPKGSYSGTYTKGDTLEFIAMTPAND
RDLKRNITQAPALNVGCTIOVGASSTPTTGKATTOADAEQVIRKQNSTVA
VASOPTIYSSSGSQSNKMLPNKKPATTVAFTVPASTTEPTVSSSTSTPIS
TWRPTEKVIETFGASEGNKIDIASKQALIAATDGVVYAGNALRGYGNLII
KNDYLSAVAHNDTMLVREQEYVAGQKIAMTGSTGSTRLHEIRIKKSVNPLR
YLPR"

protein_bind

762..777
/gene="ORF-X"
/bound_moiety="CRP"

-10_signal

797..802
/gene="ORF-X"

-35_signal

820..825
/gene="ORF-X"

protein_bind

879..894
/gene="ORF-X"

gene

/bound_moiety="CRP"

CDS

1398..1612
/gene="rpos"
1398..>1612
/gene="rpos"
/codon_start=1
/transl_table=1
/product="RNA polymerase sigma factor"
/protein_id="BA041003.1"
/db_xref="GI:4433062"
/translation="MSQNTLKVHNDINDEAEFDENGVEVDEKALVELEPSNDIAEEB
LISQATGVVDATQLYGEIGYSPLLT"

BASE COUNT

430 a 397 c 421 g 364 t

ORIGIN

Alignment Scores:

Pred. No.: 0.000832 Length: 1612
Score: 206.00 Matches: 94
Percent Similarity: 41.18% Conservative: 60
Best Local Similarity: 25.13% Mismatches: 142
Query Match: 12.34% Indels: 79
DB: 1 Gaps: 15

US-10-018-706-2 (1-322) x ECORPOS (1-1612)

QY 3 VAlThrIleAlaIleAsnSerGlnAsnGlnLysProIleLeuArgLeuGlyLeuIlePhe 22
DB 276 CTGACACTTCAAAATCACCGGACCGGTCAGCTCCGTTAATGCAATGCGCTGCAATA 335
QY 23 GlyValIleThrThrCysIleLeu----- 30
DB 336 ---CTAATTCTGTATGTTGATTACGCGCGCCGCAAAATGGGAGACGACTACAGCGC 392

QY 31 -----AlaGlyCysAlaSerLysProThr-----TyrAsnSerThrSeryl 44
DB 393 AGAACCGCAATTCAGCCGGTGCAGCAGCACCAATTCAGCTACTCAACAACCGCAAT 452
QY 45 SerGlySerHisArgThrSerylSerylGlyLeu-AlaIleGlySerGlnValIleTh 64
DB 453 CCAGCAGTGCAGCAGTACGCTCAGCAGCCGCTACAGATGAGAAAACGAGCCATGCTCA 512
QY 64 rAspSerGln-----GlyValPro-----AsnArgTyrGlnValIly 76
DB 513 TTAACGTCAGTATGGGAACATTCGAAAGGAGATTATACGGCAGTACCTTATACCGTAA 572
QY 76 sGlnGlyAspThrValSerLysIleAlaGlnArgTyrGlyLeuAsnTrpAsgGluIleG 96
DB 573 AAAGGCGACACACTTTCTATATCGCTGATTAATCTCCACAGATTCCTCGTACCTTGC 632
QY 96 yHisIleAsnLeuAsnSerSerylTyrThrIleTyrThrGlyGlnTrpLeuThrLeuTr 116
DB 633 TCAGCGCAACAATATTCAGCAGCATACGCGCTGAAAGCTGTCAG-----ACCTTCA 686
QY 116 pSerGlyAsp-LeuLysValArgGluArgSerIleSerSerylValAsnThrAlaHisT 136
DB 687 GGTGGTATATGCTTCGGTACGCCA-----TCACTGGCGGAAA 725
QY 136 hrProSerProValAlaValGlnSerSerArgProProValGlnGlnHisProAlaValG 156
DB 726 TGCATTATCC-----AGCGCAGCAGCAGCAGCAGCAAGCA- GTTGTA 766
QY 156 InLysPro-----ThrProValValValValLysProThr----- 169
DB 767 TCAAGCCTGCACAAATTCACCGTGTGCTGTGGTGCAACCGCAATTAAGTATCTG 826
QY 170 -----ProThrProPro 174
DB 827 AGCTTCGGGTGAACAGAGTCTAACAATATGTCGCAACAGCAAGCAATCGGACCA 886
QY 174 aValGlnGlnProAlaProValAlaProProValThrGluAlaProPheAlaThrGly- 193
DB 887 CGGTACACGCGCTGTACCGTACCAACAGCAGCAGCAGCAGCAGCAGCAGTCA 946
QY 194 -----SerSerGlyValMetGlnPheArgTyrProValGlyAlaThrAsnProVal 211
DB 947 CATCAACCAATACCGCTATCTCCACCTGCGCTGCGC-----ACTGAGGGCAAGTGA 1000
QY 211 aLArgArgPheGlyThrAlaThrValAlaGlySerThrValThrSerAsnGlyMetTrp 231
DB 1001 TCGAAACCTTTGGC---GCTTCTGAGGGGGC-----AACAAAGGGGATTGATA 1045
QY 231 heSerGlyArgAspGlyAspLeuIleAsnAlaSerAsnAlaGlyThrValIleGlnAla 251
DB 1046 TCGCAGGACGAAAGAGCAGCAATATTCGACCGCAGATGCGCGCTGTATTATGCTG 1105
QY 251 sPhIAsnMetAspGlyAlaSer-----IleValIleGlnHisThrAsnGlyPheValS 269
DB 1106 GTAACGCGCTCGCGGCTACGCTATCTGATTATCATTAACATATATGATATACCTGA 1165
QY 269 erSerTyrIleHisIleLysAspAlaGlnValLysThrGlyAspThrValArgThrGlyG 289
DB 1166 GTGCTACGCGCCATACAGCAACAATGTGCTGCGGGAACAACAAGATTTAAGCGGGCG 1225
QY 289 InArgIleAlaSerMetLysAsnGlnProSerGlyAlaAlaLeuPheGluPheArgIleS 309
DB 1226 AAAAATATGCAACATGGGTAGCACCGGACAGTCAACAGCTTGATTTGAAATTC 1285
QY 309 erArgAsnGlyValTyrValAspProLeuThrValLeu 321
DB 1286 GTTACAAGGGGAATTCGTTAACCCTGCTGCTATTG 1323
RESULT 39
AE009547 12362 bp DNA linear BCT 28-DEC-2001
LOCUS AE009547/c
DEFINITION Brucella melitensis strain 16M chromosome I, section 104 of 195 of

the complete sequence.
AF009547 AB008917
AE009547.1 GI:17983036

Brucella melitensis
Bacteria, Proteobacteria; alpha subdivision; Rhizobiaceae group;
Brucellaceae; Brucella.
1 (bases 1 to 12362)
Delvecchio,V.G., Kapralov,V., Redkar,R.J., Patra,G., Mujer,C.,
Los,T., Ivanova,N., Anderson,I., Bhattacharya,A., Lykidis,A.,
Reznik,G., Jablonksi,L., Larsen,N., D'Souza,M., Bernal,A.,
Mazur,M., Goldsman,E., Selkov,E., Elzer,P.H., Hagius,S.,
O'Callaghan,D., Letesson,J.-J., Haselkorn,R., Kyridides,N. and
Overbeek,R.
The genome sequence of the facultative intracellular pathogen
Brucella melitensis
Proc. Natl. Acad. Sci. U.S.A. 99 (1), 443-448 (2002)
11756688
2 (bases 1 to 12362)
Delvecchio,V.G., Redkar,R.J., Patra,G. and Mujer,C.
Direct Submission
Submitted (13-NOV-2001) Institute of Molecular Biology and
Medicine, University of Scranton, Scranton, PA 18510, USA
3 (bases 1 to 12362)
Elzer,P.H. and Hagius,S.
Direct Submission
Submitted (13-NOV-2001) Department of Veterinary Science, LSU Ag
Center, 111 Dalmonte Building, Baton Rouge, LA 70803, USA
4 (bases 1 to 12362)
Kapralov,V., Los,T., Ivanova,N., Anderson,I., Bhattacharya,A.,
Lykidis,A., Reznik,G., Jablonksi,L., Larsen,N., D'Souza,M.,
Bernal,A., Mazur,M., Goldsman,E., Selkov,E., Haselkorn,R.,
Kyridides,N. and Overbeek,R.
Direct Submission
Submitted (13-NOV-2001) Integrated Genomics, Inc., 2201 W. Campbell
Park Drive, IL 60612, USA
5 (bases 1 to 12362)
Letesson,J.-J.
Direct Submission
Submitted (13-NOV-2001) Unite de Recherche en Biologie Moleculaire,
Laboratoire d'Immunologie et de Microbiologie, Universite of Namur,
61 rue de Bruxelles, Namur 5000, Belgium
6 (bases 1 to 12362)
O'Callaghan,D.
Direct Submission
Submitted (13-NOV-2001) Faculte de Medecine, INSERM U431, Avenue
Kennedy, Nimes 30900, France

1. 12362
/organism="Brucella melitensis"
/strain="16M"
/db_xref="taxon:29459"
/chromosome="1"
/complement(108..1563)
/gene="BME11069..1563"
/complement(106..1563)
/gene="BME11069"
/EC_number="5.2.1.8"
/codon_start=1
/transl_table=11
/product="TRIGGER FACTOR, PPIASE"
/protein_id="AAU52250.1"
/db_xref="GI:17983037"
/translation="MTRSEGLMNVETELNEGLKEIKVVPAGDLLEAKLARLETAR
GARINGFPGVPTAHLKMYGKSFMAIINEIDSSRSILARNEKSAVOPVIM
SDEKEAEKVLGKADPVLNVEVLPALEVDPSKIAVTRVVDISDEVDQVARI
ASTRTPTETKGAENEDRVITIDYLGKIDPEFGADNDQVLVSGQFIDGFEOL
IGLKAGDEKIVITVTPAEYGAHLAKGKATPDIKYEVAKPNELVLDDETKAGLIES
LERLRQVREOIESOYGOITROKVKQOILDALDGDVQFETPOKLVDABENNIWOQINF
DIOGRVTEDETTTEARERERKLAEKRRVLGLVLSIGKAGVETVEELQAVY
DQVRVPGQEKRTYDILRTPTPAVNLRIKPIPEKVVHLLANINVTDKVSKELTA
EBEDASERKPAKKAARKKAPKKAEGSGBEA"

gene
CDS
tRNA
gene
CDS
complement(1627..1709)
/product="tRNA-Leu"
/complement(1943..2512)
/gene="BME11070"
/complement(1943..2512)
/gene="BME11070"
/codon_start=1
/transl_table=11
/product="Hypothetical Protein"
/protein_id="AAU52251.1"
/db_xref="GI:17983038"
/translation="MTNKKWLYAGAVLAAALLOTGITLYAGIEKRAALIRSGODIVPOT
EPVDPDLWRGDVVVIGYEISNIRSAIGGVPRAGSRVYVAVKPPAKISRRSASF
VPEKDLATGERTGAGYEISDPPANIRLSFGIERRYVVEGEGHPIGSGQREHDIT
VYAVYASGPVYKALMDGQPIREGPLV"
complement(2509..3513)
/gene="BME11071"
complement(2809..3513)
/gene="BME11071"
/codon_start=1
/transl_table=11
/product="Hypothetical Membrane Spanning Protein"
/protein_id="AAU52252.1"
/db_xref="GI:17983039"
/translation="MATLGGLLGAAVIMVLAANQEMPLRIGVYFVLIWASYIGG
AMRQAGDKVFPALVYLGASFGAGIALVGOYTHISGDVYSALWTYTLVASFLA
RAQALAFGAGVACFYLTSPFADSNLSGADISRYWGPDLILAGAAALFTSRHAA
HFLALFSIGCLILVAGQENKTVLIMTVIGLILADGIRHQLQKTRFAPHLAY
GLILVLSFALILDVITIGVGSAGIDRIDILYSMLILALISGAIAICGRDNGLRSI
AYAAFSIEVLYLAETVGTIGTGFELTGLIIVLILAAFVRMESRFRKQGLEAPH
A"
3919..4065
/gene="BME11072"
3919..4065
/gene="BME11072"
/codon_start=1
/transl_table=11
/product="Hypothetical Protein"
/protein_id="AAU52253.1"
/db_xref="GI:17983040"
/translation="MNLFRSYNNWRRRTVETVNLNQLSTRELDNDIGISRADIPVAVQ
AKAR"
4271..5707
/gene="BME11073"
4271..5707
/gene="BME11073"
/codon_start=1
/transl_table=11
/product="GLUCOSE INHIBITED DIVISION PROTEIN A"
/protein_id="AAU52254.1"
/db_xref="GI:17983041"
/translation="MKALIKQDKGIAANSNTDLSPPVYIGGGLAGSEANQIAGVAP
VVLHEPVRGTDAHKTQQLAEVCSNPSFSDATNAVVLHAEMKLAOSLIACADP
AHQVPAGGALAVDEGFSQAVTALAEHPIITIEREITGLPTPEWGTITIIATGPLA
PSLAEALAEETDADALAFDPAIPIIHFDISINDVCFQSVMERGVPETLRHGKMGKINP
MDKQYAFVAALEGGKTDPEKEWEGPPEYDGLPIVMAERGGPPTLRHGKMGKINP
NHNRPVYKPAVVOLEODNALGTLNNVWGFGTKLKYSQSGIAPFMIIGLNAEFAKLG
GLHNRTYNSPVLIDNVLRLKSRQTLRFACQVGCSEYVSSAIGLAGRTAAERKS
QAAYPPPTTARFGLIGHTIGGHTVTDDEGKRSFGPMNVNFGLPFVDVYKPKGKSL
RGKRTIARRKALSARALADCRNMLSY"
complement(5759..6592)
/gene="BME11074"
complement(5759..6592)
/gene="BME11074"
/EC_number="2.5.1.-"
/codon_start=1
/transl_table=11
/product="PHYTOENE SYNTHASE"
/protein_id="AAU52255.1"
/db_xref="GI:17983042"
/translation="MNQETICLLALREADRRLYSVLVAPQEKRGALAAIYAPNAEI
ARIDLVIEPILPGEVRLQWNRDILNGERKSGAENAPVAAALITETIEKLEPRLAPNDY
CEARVFDLYDDPMSRNDLEGCYCGETASALITQLAGFILDRAAQAHAETAGHAGVAA

gene
complement (6598) . (6987)
/gene="BME11075"
complement (6598) . (6987)
/gene="BME11075"
/codon_start=1
/transl_table=11
/product="Hypothetical Cytosolic Protein"
/protein_id="AAL52255.1"
/db_xref="GI:17983045"
/translation="MARGIEIRAEHPGRADIDAYNGGFRFAEMSHRSILCPGSI
HGMPEKPPILSRADIGLIEASDIEILVGTGMDLRIPEDVRALLRHHISSDPM
STGAAYRTYVNLAEADRAVAALIAVD"
complement (6990) . (9317)
/gene="BME11076"
complement (6990) . (9317)
/gene="BME11076"
/codon_start=1
/transl_table=11
/product="PROTEIN TRANSLOCASE SUBUNIT SECD / PROTEIN
TRANSLOCASE SUBUNIT SECD"
/protein_id="AAL52257.1"
/db_xref="GI:17983044"
/translation="MGELOQCFIOAETDLSMLYFSRWKSAIMLAVLSLISPNP
LSRETLNLPDLPRKQVSLGLSGSRLLIYONMKTDLTTANIMRRLBELGY
GNPYEGRNQIIEVDPGLYDAQLMDILTRNLSRAMPDMSDDARCPPLAD
SEIYSGDDPPVGLTKTFLTGHDITDAKSTISADGQVITLTLDDNRRLADI
TAQGENSFALVDNVQVSAFVSPDLSLQIEAGDQVITLTLDDNRRLADI
VTVLEERTIASLGEDEYASAVLALALAVGLVPMVSLGILVIALVAVNI ILL
TAVLSLIGASISLAIAGLVLTIGLADVAHLYIYERVEDRKGYSVQVAMESEFYA
LSTVDNLTTLLAALVFLGSGVGFALVAIGITLFTTLTFRLLIAQWRT
AKPEVEVRRLKIVPTVHTIPMRLQVPTLGI SYLACIYVALFVNIGVNDIPRG
SMVELQKNGDANLEEDINERLAEINISARVLPKSPRSALVTIGSGVEDDAQIVA
VLKSGEEDYSQFQVAVVQVTSQUSRAGVAVLISLIGIFLYWFRFWMQLA
VLSLTHDVVLSGMFIVRFMEFNMVSAVLIIGISLNDVIVYDRRENRLRYKKA
PLPAIDIASINQTLSTFTFLAVPLAVPFAFGSGSRIMFALASGITIVASYSI
PIAALIVQFLKRETDAGDAVDALASLNTLS"
complement (9437) . (9799)
/gene="BME11077"
complement (9437) . (9799)
/gene="BME11077"
/codon_start=1
/transl_table=11
/product="Hypothetical Exported Protein"
/protein_id="AAL52258.1"
/db_xref="GI:17983045"
/translation="MTSEEFIMFTVPAFAQSGSVGPDMLMSILPFLIFVIVYFLI
IRPORTOMKROEMLSVRRGDVTVTGGI VGVKLKVVDNELELEIADGVRIVVRA
TLMVRYKGEFVANAANKK"
9902 . 10804

gene

Alignment Scores:

Pred. No.: 0.00902 Length: 12362
Score: 205.00 Matches: 79
Percent Similarity: 41.06% Conservative: 45
Best Local Similarity: 26.16% Mismatches: 132
Query Match: 12.28% Indels: 46
DB: 1 Gaps: 12

US-10-018-706-2 (1-322) x AB009547 (1-12362)

Qy 30 LeuAlaGlyCysAlaSerLysProThyTrpAsnSerThrSerGlySerGlySerHisArg 49
Db 11653 ATCGCTGGCGTTCCGCAACGCTGCTCAACGCAAGAAATCCTCCGCAACCAATATG 11594
Qy 50 ThrSerGlySerGlyGlyLeuAlaIleGlySerGlnValIleThnAspSerGlnGlyVal 69
Db 11593 GCATCGCTGGCGGCGGCTATGCCA-----ACCCGCGCATCATG 11555
Qy 70 ProAsnArgTrgInValIleGlnGlyAspThrValSerLysIleAlaGlnArgTrgIly 89

Db 11554 GCGCGTCCCTAAGCTGTGAAGAGCGGGAGTTCGCTGTTCTCCATCGCAAGAACACAAAT 11495
Qy 90 LeuAsnTrpArgGluIleGlyHisIleAsnAsnLeuAsnSerSerTrpTrpIleTrpThr 109
Db 11494 GTCCCTGTCCAGCAGTTGAAGCGCGCAACGCGCTGACCACTGCGC---GCCATTCGCGTC 11438
Qy 110 GlyGlnTrpLeuTrpLeuTrpSerGlyAspLeuLysValArgGluArgSerIleSerSer 129
Db 11437 GGGCAGTCGCTCGTATTT-----CCGTCGTGCGCGCA 11405
Qy 130 GlyValAsnThr-----AlaHisThrProSerProVal---AlaValGlnSerSer 145
Db 11404 GGTGCGCGGACACAGCGTCCGCAATACCGCGGACCGGCAACCGCGCAAGCGCC 11345
Qy 146 ArgProProValGlnGlnHisProAlaValGlnLysProThrProProValValVal 165
Db 11344 AGTCCGCGCGCAATGAGAGGTGCAAGAGTTAAAGCTTATACCGCGCGCAGCGCAAC 11285
Qy 166 LysLysProThrProThrProProValValGlnLysProAlaProValAlaProProVal 185
Db 11284 AAG-----GTAATCGAGATCGCGAA----- 11264
Qy 186 ThrGluAlaProPheAlaThrGlySerSerGlyValMetGlnPheArgTrpProValGly 205
Db 11263 AAGCATCAGCGCGCAGCGCATCTCAACCGCGCATCTCGCAGATGCGTGGCCGTT--- 11207
Qy 206 AlaThrAsnProValValArgArgPheGlyThrAlaThrValAlaGlySerThrValThr 225
Db 11206 ---CGTGGCGCGCATTCGCGCAAGCTTCGCGCAGCGTGA---GGCAGCTTCGCTC 11159
Qy 226 SerAsnLysMetTrpPheSerGlyArgAspGlyAspLeuIleAsnAlaSerAsnAlaGly 245
Db 11158 AGCAGCGTATCGATATCATCATGCTCGCGAAGCAGCGCGTCAAGCGCGCAAGCGGT 11099
Qy 246 ThrValIleGlnAlaAspHisAsnMetAsp-----GlyAlaSerIleValIleGlnHis 263
Db 11098 GTTGCTGATCATGCGCGCGCATGTTGTAAGGAATTTGCGCAGCGCTTCATCCGCGCAC 11039
Qy 264 ThrAsnGlyPheValSerSerTrpIleHisIleLysAspAlaGlnValLysThrGlyAsp 283
Db 11038 GACATGCGCTTGACCGTCTATGCGCATTAACGCAATATGATGTCAGCGCGCGCAG 10979
Qy 284 ThrValArgThrGlyGlnArgIleAlaSerMetLysAsnGlnProSerGlyAlaAla--- 302
Db 10978 AAGTTCCCGTGGCGCAAGAAATGCC-----AAGTCGCGCATGACCGCGCAAGCGCAAG 10925
Qy 303 -----LeuPheGluPheArgIleSerArgAsnGlyValIleValAspProLeuThrVal 320
Db 10924 TCGCGCAAGCTGCACTTCGAGGTGCGCAAGAACTCGCGCGTCAATCGCAACAAATAT 10865
Qy 321 LeuLys 322
Db 10864 CTGGAA 10859

RESULT 40
NMA522491/c 329861 bp DNA linear BCT 04-DEC-2000
LOCUS
DEFINITION
Neisseria meningitidis serogroup A strain Z2491 complete genome;
segment 5/7.
VERSION
AL162756.2 GI:7380091
KEYWORDS
SOURCE
Neisseria meningitidis Z2491.
Neisseria meningitidis Z2491.
Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
Neisseria.
REFERENCE
1 (bases 1 to 329861)
Parkhill,J., Achtman,M., James,K.D., Bentley,S.D., Churcher,C.,
Klee,S.R., Morelli,G., Basham,D., Brown,D., Chillingworth,T.,
Davies,R.M., Davis,P., Devlin,K., Feltham,T., Hamlin,N.,
Holroyd,S., Jagsels,K., Leather,S., Moutie,S., Mungall,K.,
O'Neill,M.A., Rajandream,M.A., Rutherford,K.M., Simmonds,M.,
Skellern,D., Whitehead,S., Spratt,B.G. and Barrell,B.G.

```

TITLE      Complete DNA sequence of a serogroup A strain of Neisseria
            meningitidis Z2491
JOURNAL    Nature 404 (6777), 502-506 (2000)
MEDLINE    20222556
PUBMED     10761919
REFERENCE  2 (bases 1 to 329861)
AUTHORS    Parthi11,T
TITLE      Direct Submission
            Submitted (30-Mar-2000) Submitted on behalf of the Neisseria
            sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
            Hinxton, Cambridge CB10 1SA E-mail: parthi11@sanger.ac.uk
COMMENTS   Notes:
            Details of N. meningitidis sequencing at the Sanger Centre are
            available on the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/N_meningitidis/).
FEATURES   source
            1..329861
            /organism="Neisseria meningitidis Z2491"
            /strain="Z2491"
            /db_xref="taxon:122587"
            /note="serogroup: A"
            194..565
            /gene="NMA1439"
            194..565
            /gene="NMA1439"
            /note="NMA1439, possible lipoprotein, len: 123 aa;
            unknown, contains a probable N-terminal signal sequence
            and an appropriately positioned P800013 prokaryotic
            membrane lipoprotein lipid attachment site"
            /codon_start=1
            /transl_table=11
            /product="putative lipoprotein"
            /protein_id="CAB84675.1"
            /db_xref="GI:7380092"
            /db_xref="SPTREMBL:O9JU92"
            /translation="MAKTIISIPVALILGGCAAGGNPFGSLDGTGKGSIVKAVK
            SCRRALNKRSEMRITPALMSAKQEMENKICACVADAPQLTGNVDMQMLDSETR
            NOALALITAKTYSACFKHLYR"
            212..244
            /gene="NMA1439"
            /note="P800013 prokaryotic membrane lipoprotein lipid
            attachment site"
            complement(571..580)
            /note="Core DNA uptake sequence: gccgctcgaa"
            /label=DUS
            592..601
            /note="Core DNA uptake sequence: gccgctcgaa"
            /label=DUS
            638..2047
            /gene="chrc"
            638..2047
            /gene="chrc"
            /EC_number="4.2.99.2"
            /note="NMA1440, thrC, probable threonine synthase, len:
            485 aa; similar to many e.g. SW:THRC_METG (EMBL:D14071),
            thrC, Methylobacillus glycosenes threonine synthase (EC
            4.2.99.2) (475 aa), fasta scores: E(): 0, 61.9% identity
            in 475 aa overlap. Contains Pfam match to entry PF00291
            S.T. dehydrogenase, Pyridoxal-phosphate dependent enzymes"
            /codon_start=1
            /transl_table=11
            /product="putative threonine synthase"
            /protein_id="CAB84676.1"
            /db_xref="GI:7380093"
            /db_xref="SPTREMBL:O9JU91"
            /translation="MKYISTRGETLHAKPSESVILMGLAPDGLMPEHYPOIGREALD
            KMRGLAYPELAEIMRLPVTDIPEDDLRLNRYTEAFGKEITPVRTLSDGKIQ
            ALSNGPTLAKOMAFLOFGNAEYVNLNKGKLTILGATSGDGSAAEVALGKKGIVN
            FVMSPDGKMSAFORAOYMSILOGDGNININAVKGMFDDCDIVKAOVNAAPAEKYGIG
            TVNSINWGRIVAOVYVYFAGYFKATQSDNDEQVSCVPSGNFNVCAGHIAKMGGLPVR
            RLIVATNENDIVLDEFFKTYGAYRPNSAHYVSSPMDSKSNFERFVFDLMRDPQ
            EINTLMAEVAAGKGNLRPALDKVGKGFSGKSTHADRLLTIROYVEODELIDPH
            TADGVKVAEREVERETVCTETALAAKADATIRBAVGVAIPRPAALGEGENLPORV

```

```

misc_feature 875..1780
            /gene="chrc"
            /note="Pfam match to entry PF00291 S.T. dehydrogenase,
            Pyridoxal-phosphate dependent enzymes, score 231.80,
            E-value 1e-65"
            complement(1922..931)
            /note="Core DNA uptake sequence: gccgctcgaa"
            /label=DUS
            2094..2888
            /gene="NMA1441"
            2094..2888
            /gene="NMA1441"
            /note="NMA1441, len: 264 aa; unknown"
            /codon_start=1
            /transl_table=11
            /product="hypothetical protein NMA1441"
            /protein_id="CAB84677.1"
            /db_xref="GI:7380094"
            /db_xref="SPTREMBL:O9JU90"
            /translation="MTVMQETPSTLIDELAAKQEAALAPHLADGTXYWRKAGRH
            NAKWRVALGVAVRYLKIYVLPVPSGGEPAITESKRLYEALAGIAVPELAVRK
            NALMGNLGEGILDTQIROEAEGKADAMLAGLEAIRVHKQOFLSQAFARMMMSDG
            KNISFLDEDDPSEVLTIACQADMDICYHTSLILKNGSLLEAAAEKNGVLSDDP
            AEIOKLIATVYKPIIPRLRLEHPRWGDAIRLAISILSLADMP"
            2904..2937
            /note="Stem loop containing DNA uptake sequences: acgat
            gccgctcgaa gccgctcgaa atata"
            2909..2918
            /note="Core DNA uptake sequence: gccgctcgaa"
            /label=DUS
            complement(12923..2932)
            /note="Core DNA uptake sequence: gccgctcgaa"
            /label=DUS
            3065..3066
            /label=DUS
            3076..3082
            /gene="fpr"
            3076..3082
            /gene="fpr"
            /EC_number="1.18.1.2"
            /note="NMA1442, fpr, probable ferredoxin-NADP reductase,
            len: 258 aa; similar to many e.g. SW:FENK_AZOVI
            (EMBL:L36319), fpr, Azotobacter vinelandii
            ferredoxin-NADP reductase (EC 1.18.1.2) (257 aa), fasta
            scores: E(): 0, 69.8% identity in 255 aa overlap. Similar
            to NMA1664, fasta scores: E(): 1.3e-30, 34.6% identity in
            257 aa overlap. Contains Pfam match to entry PF00175
            oxidored fad, Oxidoreductase FAD/NAD-binding domain"
            /codon_start=1
            /transl_table=11
            /product="putative ferredoxin-NADP reductase"
            /protein_id="CAB84678.1"
            /db_xref="GI:7380095"
            /db_xref="SPTREMBL:O9JRE3"
            /translation="MAAFNTOKLVSHVHMWDAYPTPTCTRDESIFRNGGFVWGLMV
            DGRGLMAYGASANMWEHSEFISIKVODGPTSLQLHLKVGDDVLSKPTGTLVAG
            DINNGKLVLLSTGTAPLSITKDEIYEOGFSLIKVGVKRLAYADRTKEL
            BEHYLDIVKELIKIYPIVSREREHGHSLDLMVSGKLFEDIGLPIKINPDORAML
            CGSRAMKDKRCVKVDRLGLVSPRTVGRGYLLERAVVDQ"
            3394..3766
            /gene="fpr"
            /note="Pfam match to entry PF00175 oxidored fad,
            Oxidoreductase FAD/NAD-binding domain, score 20.00,
            E-value 8.4e-05"
            3866..3875
            /note="Core DNA uptake sequence: gccgctcgaa"
            /label=DUS
            3964..3983
            /note="Core DNA uptake sequence: gccgctcgaa"
            /label=DUS
            complement(3984..4038)
            /note="RS103"
            /label=RS103

```

```

repeat_unit 4039..4058
             /note=">= 90% match to ATTCNNNNNNNNNGGAAT"
             /label=DRS3
             /gene="NMA1443"
             /db_xref="GI:7380096"
             /db_xref="SPTREMBL:Q9JUB9"
             /translacion="MIYKNCNPKCAENCKMVFLENGSATVIRGIMKKIIT
             IDIITPFA"
             /complement(4509..6980)
             /gene="NMA1444"
             /complement(4509..6980)
             /note="NMA1444, probable P-type cation-transporting
             ATPase, len: 823 aa; similar to many e.g. SW:COXA_ENTHR
             (EMBL:L13292), COXA, Enterococcus hirae
             coper/potassium-transporting ATPase A (EC 3.6.1.36) (727
             aa), fasta scores; E(): 0, 34.0% identity in 744 aa
             overlap. Similar to NMA1539, fasta scores; E(): 0, 34.4%
             identity in 735 aa overlap. Contains hydrophobic, probable
             membrane-spanning regions. Contains two Pfam matches to
             entry PF00122 E1-E2_ATPase, E1-E2_ATPases and P500154
             E1-E2_ATPases phosphorylation site"
             /codon_start=1
             /transl_table=11
             /product="putative P-type cation-transporting ATPase"
             /protein_id="CAB84680.1"
             /db_xref="GI:7380097"
             /db_xref="SPTREMBL:Q9JUB8"

```

Alignment Scores:

Pred. No.:	0.444	Length:	329861
Score:	203.00	Matches:	71
Percent Similarity:	43.46%	Conservative:	42
Best Local Similarity:	27.31%	Mismatches:	97
Query Match:	12.16%	Indels:	50
DB:	1	Gaps:	11

US-10-018-706-2 (1-322) x NMA522491 (1-329861)

```

Qy 73 TyrGlnValIysGlnGlyAspThrValSerIysIleAlaGlnArgTyrGlyLeuAsnTrp 92
Db 283427 CATNCTANTGTCGCGCGGACGCGTGTACACATTTCCAAACGCTACCATATCTCTCAA 283368
Qy 93 ArgGluIleGlyHisIleAsnAsnLeuAsnSerSerTyrThrIleTyrThrGlyGlnTrp 112
Db 283367 GACGATTTCCGTCGCGTGAACGCGATGACCGACAAT--ACGTTGAGCATCGCTCAG--- 283314
Qy 113 LeuThrLeuTrpSerGlyAspLeuIysValArgGluArgSerIleSerSerGlyValAsn 132
Db 283313 -----ATTGTTAAAGTCAAAACGCGAGATATGCGCA----- 283281
Qy 133 ThrAlaHisThrProSerProValAlaValAlaGlnSerSerArgProProValGlnGlnHis 152
Db 283280 -----CCGAAGCCGCGACCGCTTAAA--AGCAGCGCCGCGCTACCG----- 283242
Qy 153 ProAlaValGlnIysProThrProProValValValValValValValValValValValPro 172
Db 283241 GCTGCGCGCAACCCCTGTG-----AAACCGCGCGCAACCG 283203
Qy 173 ProValValGlnGlnProAlaProValAlaPro----- 183
Db 283202 CCTGTGCGAGTCCGCGCGCAACCTGCGCGCCGCTGCGGAAATAAAGCGGTTCCGCGC 283143
Qy 184 ProValThrGluAlaProPheAla-----ThrGlySerSerGlyValMetGln 199

```

```

Db 283142 CCCCCTCCGCAATCTCTGCGCTTCCCTCCGCGACGCGTTCGTCGCGCATTTGT 283083
Qy 200 PheArgTyrProValGlyAlaThrAsnProValValArgTrpPheGlyThrAlaThrVal 219
Db 283082 TGGCAGCGTCCG-----ACGCAAGTAAAGTGTTCCTCCGATTTCCGCGC----- 283038
Qy 220 AlaGlySerThrValThrSerAsnGlyMetTrpPheSerGlyArgAspGlyAspLeuIle 239
Db 283037 -----AACAAACAAGCGTGTGATTTTCAGGAATAAGGGGACACCCGTT 282993
Qy 240 AsnAlaSerAsnAlaGlyThrValIleGlnAlaAspHisAsnMetAspGlyAlaSer-- 258
Db 282992 TTGCGCGCGCTGACGCGCAAAAGTGTATTATGACAGTTCCGATTTGAGGGGATACGCCAAT 282933
Qy 259 ---IleValIleGlnHisThrAsnGlyPheValSerSerTyrIleHisIleValAspAla 277
Db 282932 TTGTCATCATTCGACGCAATATTTCTTCTCTGACCGCATPACGGCAACAAATAATG 282873
Qy 278 GlnValIysThrGlyAspThrValArgThrGlyGlnArgIleAlaSerMetIysAsnGln 297
Db 282872 CTGTCGCGCAAGCCAGCAGGTAAACGCGGACGACAGTCTTGATGCGCAATACC 282813
Qy 298 ProSerGlyAlaAlaLeuPheGluPheArgIleSerArgAsnGlyValTyrValAspPro 317
Db 282812 GAGGCTTGAACAGCAGCTTCATTTGAGGTGCGCAAAACGCAACCGTTAATCCG 282753

```

Search completed: July 3, 2003, 21:18:47
 Job time : 2677 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 3, 2003, 13:40:17 ; Search time 198 Seconds

(without alignments)
3662.341 Million cell updates/sec

Title: US-10-018-706-2

Perfect score: 1670
Sequence: 1 MTVTAINSONKPKIRLGL.....LFEFRISRNQVYDPLTVLK 322

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2185239 segs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seg length: 0
Maximum DB seg length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODE=frame_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US10018706/runat.30062003_091104_23784/app_query.fasta_1.519
-DB=N.GeneSeq.101002 -OPMT=fastcap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blonsum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=40
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MTLEN=0 -MAXLEN=200000000
-USER=US10018706@cgn2_1.1.200@runat.30062003_091104_23784 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBPOOLC=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N.GeneSeq.101002.*

1: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/NA1980.DAT.*
2: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/NA1981.DAT.*
3: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/NA1982.DAT.*
4: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/NA1983.DAT.*
5: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/NA1984.DAT.*
6: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/NA1985.DAT.*
7: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/NA1986.DAT.*
8: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/NA1987.DAT.*
9: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/NA1988.DAT.*
10: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/NA1989.DAT.*
11: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/NA1990.DAT.*
12: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/NA1991.DAT.*
13: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/NA1992.DAT.*
14: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/NA1993.DAT.*
15: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/NA1994.DAT.*
16: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/NA1995.DAT.*
17: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/NA1996.DAT.*
18: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/NA1997.DAT.*
19: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/NA1998.DAT.*
20: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/NA1999.DAT.*
21: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/NA2000.DAT.*
22: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/NA2001A.DAT.*
23: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/NA2001B.DAT.*
24: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1670	100.0	969	22 AAF30046	Moraxella catarrhalis
2	1670	100.0	48328	22 AAF28540	Genomic fragment #
3	1666	99.8	966	22 AAF30047	Moraxella catarrhalis
4	219.5	13.1	1830121	17 AAF242063	Haemophilus influenzae
5	209	12.5	2179	21 AAF93349	Haemophilus somnus
6	200	12.2	1248	21 AAF53075	Neisseria meningitidis
7	200	12.0	1014	21 AAF53073	Neisseria meningitidis
8	198	11.9	1056	21 AAF53074	Neisseria meningitidis
9	198	11.9	6100	21 AAF81766	Neisseria meningitidis
10	198	11.9	349980	21 AAF21610	Neisseria meningitidis
11	198	11.9	349980	21 AAF21611	Neisseria meningitidis
12	198	11.9	837096	21 AAF81489	N. meningitidis pa
13	165	9.9	1885	14 AAF51084	H. somnus lipb gen
14	158.5	9.5	684707	24 ABO67196	Listeria innocua c
15	158.5	9.5	2944528	24 ABA03041	Listeria monocytogenes
16	158.5	9.5	3011208	24 ABO68245	Listeria innocua d
17	147	8.8	2365589	24 ABA90521	Genomic sequence o
18	146.5	8.8	615	22 AAF67706	C glutamicum codin
19	146.5	8.8	349980	22 AAF68532	Actinobacillus act
20	145	8.7	949	22 AAF59421	Polynucleotide seq
21	140	8.4	14063	20 AAF20500	Listeria innocua c
22	139.5	8.4	1163020	24 ABO67197	Haemophilus influe
23	135	8.1	1830121	17 AAF242063	H influenzae BASB2
24	134	8.0	1233	24 AAF46624	Plasmid PCR28 whi
25	131	7.8	1233	24 AAF46624	Lysein coding seque
26	131	7.8	3646	14 AAF51086	CDNA encoding huma
27	130.5	7.8	1929	17 AAF06135	Micromonospora car
28	130.5	7.8	4093	21 AAF23454	Streptomyces venez
29	130.5	7.8	11115	23 AAF50562	S. venezuelae vep
30	130.5	7.8	15872	18 AAF68715	Genomic sequence o
31	130.5	7.8	15872	21 AAF68715	Mycobacterium tube
32	130.5	7.8	2365589	24 ABA90521	H influenzae BASB2
33	130.5	7.8	4403765	22 AAF99682	Achromobacter lyti
34	130.5	7.8	4411529	22 AAF99682	Borrelia burgdorfe
35	130	7.7	1233	24 AAF46625	Borrelia burgdorfe
36	128.5	7.7	1520	21 AAF49721	Beta-lytic protease
37	128.5	7.7	111309	20 AAF20250	HEV-US2 full lengt
38	128.5	7.7	910715	20 AAF20248	Arabidopsis thalia
39	127.5	7.6	1520	13 AAF25083	Human cervical can
40	127	7.6	7277	20 AAF20266	Nucleic acid seque
41	127	7.6	7277	20 AAF20267	Human CDNA seque
42	125.5	7.5	942	24 AAF98586	
43	125.5	7.5	1536	22 AAF73004	
44	125	7.5	903	20 AAF200243	
45	124.5	7.5	3004	22 AAF15779	

ALIGNMENTS

RESULT 1	AAF30046	standard; DNA; 969 BP.
ID	AAF30046	
XX	AAF30046	
AC	AAF30046	
XX	AAF30046	
DT	23-APR-2001	(first entry)
XX	23-APR-2001	
DE	Moraxella catarrhalis DNA encoding BASB10 protein.	
KW	BASB10; infection; otitis media; pneumonia; gene therapy;	
KW	diagnosis; antibacterial; antimicrobial; genetic immunisation;	
KW	vaccine; ss.	
XX		
OS	Moraxella catarrhalis.	
XX		
Key	Location/Qualifiers	
FT	conflict	replace(310,G)

/*tag= a
/note= "gene sequence differs from PCR product
of AAF30047 at this base"

WO200100838-A1.

04-JAN-2001.

23-JUN-2000; 2000WO-EP05854.

25-JUN-1999; 99GB-0015031.

(SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

Thomard J;

WPI: 2001-112459/12.

P-PDB; AAB20106.

Novel BASB10 polypeptides of Moraxella catarrhalis, useful as a vaccine for treating Moraxella catarrhalis infections -

Claim 10; Page 81; 88pp; English.

The present sequence is that of the coding region of the BASB10 gene of Moraxella catarrhalis Mc2931 (ATCC 43617), a causative agent of otitis media in children and pneumonia in adults. The gene encodes the 322-amino acid BASB10 protein (see AAB20106). It differs from a BASB10 DNA PCR amplification product (see AAF30047) by having A at position 310 rather than G. This would cause an amino acid change from Ser-104 to Gly in the translated sequence. The invention provides BASB10 polypeptides, and polynucleotides encoding them, as well as expression vectors, host cells and methods for producing BASB10 polypeptides using recombinant methods. Also claimed is a vaccine composition comprising a BASB10 polypeptide, an immunogenic fragment of a BASB10 polypeptide, or a polypeptide having at least 85% amino acid sequence identity to BASB10, or comprising a polynucleotide encoding such a polypeptide. A claimed method of diagnosing a Moraxella infection involves identifying a BASB10 polypeptide or antibody. A claimed therapeutic composition useful in treating humans with M. catarrhalis infection comprises at least 1 antibody directed against a BASB10 polypeptide. CC polynucleotides also have utility in diagnosis of the strain and CC type of infection, and also for therapeutic or prophylactic purposes, in particular genetic immunisation.

Sequence 969 BP; 265 A; 214 C; 242 G; 248 T; 0 other;

Alignment Scores:

Pred. No.: 1-876-116 Length: 969
Score: 1670.00 Matches: 322
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0

US-10-018-706-2 (1-322) x AAF30046 (1-969)

QY 1 MetThrValThrIleAlaIleAsnSerGlnAsnGlnuysProIleLysArgLeuGlyLeu 20
DB 1 ATGACTGAGCAGATGACATCAATTCACAAATCAAAACCATCAAGCATGGGCTTG 60
QY 21 IlePheGlyValIleThrThrCysIleLeuAlaGlyCysAlaSerLysProThrTyrAsn 40
DB 61 ATTTTGTGTGATCACCACCTTGCAATTTGGCAGATGTGCCAGTAAAGCCAACTAAT 120
QY 41 SerThrSerGlySerGlySerHisArgThrSerGlySerGlyGlyLeuAlaIleGlySer 60
DB 121 AGTACCTCAGGTTGGGAGATCATCTTCAAGTTCAAGTGGTTTGGCAATAGTTCA 180
QY 61 GlnValIleThrAspSerGlnGlyValProAsnArgTyrGlnValGlnGlnGlyAspThr 80
DB 181 CAGGTATCAGGACATCATGAGGCGTACAAATGCTATCATGATGAGCAAGGCGACT 240

QY 81 ValSerLysIleAlaGlnArgTyrGlyLeuAsnTrpArgGluIleGlyHisIleAsnAsn 100
DB 241 GTCAGTAAATGCTCAGCGCTAATGATGAATTAATTTGGCGTAAATGACACATTAATAT 300
QY 101 LeuAsnSerSerTyrThrIleTyrThrGlyGlnTrpLeuThrLeuTrpSerGlyAspLeu 120
DB 301 CTAAATAGCAGTTAAAGATTTATACAGGTCATAGCGTCACTTATAGTACAGGATCTC 360
QY 121 LysValArgGluArgSerIleSerSerGlyValAsnThrAlaHisThrProSerProVal 140
DB 361 AAGGTGGGTAGCGCTATATCATCGCTGGGTGAATACACTCAACACCTTCCCTGTG 420
QY 141 AlaValGlnSerSerArgProProValGlnGlnHisProAlaValGlnLysProThrPro 160
DB 421 GCGGTTCAGTCAGCAGACACACAGTACAGCAGACATCTCCGTACAAACCCACGCCA 480
QY 161 ProValValValValLysLysProThrProThrProProValValGlnGlnProAlaPro 180
DB 481 CTTGTGTGTGTGTAATAAAACCCACACGACTCCGCTGTGTGTTAGAGCAGCACCA 540
QY 181 ValAlaProProValThrGluAlaProPheAlaThrGlySerSerGlyValMetGlnPhe 200
DB 541 GTTGCCCCACAGTACAGACAGACACATTTGCCACGGTATGCTCAGGGGTGATGCAATT 600
QY 201 ArgTyrProValGlyValThrAsnProValValArgArgPheGlyThrAlaThrValAla 220
DB 601 CGGTATCTGTGTGTGACCAATCCAGTGTTCACGCTTGTGTGTGTGTGTGTGTGTGT 660
QY 221 GlySerThrValThrSerAsnGlyMetTrpPheSerGlyArgAspGlyLysLeuIleAsn 240
DB 661 GCGTCACTTACAGTAAATGACATGTGTTTTCGACAGATATGCAATTAATTAATAC 720
QY 241 AlSerAsnAlaLysThrValIleGlnAlaAspHisAsnMetAspGlyAlaSerIleVal 260
DB 721 GCGAGTAAACAGACAGACATTCATCAAGCTGATCAAAATATGACCGGCGAGTATGTG 780
QY 261 IleGlnHisThrAsnGlyPheValSerSerTyrIleHisIleLysAspAlaGlnValLys 280
DB 781 ATTCAGATACCAATGATGATTTGTTCAAGCTATATTCATATTAAGACCTCAAGTTAAA 840
QY 281 ThrGlyAspThrValArgThrGlyGlnArgIleAlaSerMetLysAsnGlnProSerGly 300
DB 841 ACGGGGATACCGGTGTGTCGCTCAGCCTATTGCAAGCATGAAAAATCAGCAAGCGGT 900
QY 301 AlaAlaLeuPheGlnPheArgIleSerArgAsnGlyValTyrValAspProLeuThrVal 320
DB 901 GCGGCACTATTGAAATTTGAATTTCTAGAAATGCGGTATGTTGATTCATTGACAGTA 960
QY 321 LeuLys 322
DB 961 CTYAAA 966
RESULT 2
AAF28540
ID AAF28540 standard; DNA; 48328 BP.
XX
AC AAF28540;
XX
DT 04-APR-2001 (first entry)
XX
DE Genomic fragment #27.
XX
KW Genomic library; bacteria; human upper airway; otitis media; sinusitis;
XX bronchopulmonary; endocarditis; meningitis; ss.
XX
OS Moraxella catarrhalis.
XX
PN WO200078968-A2.
XX
PD 28-DEC-2000.
XX
PF 16-JUN-2000; 2000WO-US16649.

XX 18-JUN-1999; 99US-0140121.
 PR (INCY-) INCYTE GENOMICS INC.
 PA Lagace RE, Patterson C, Berg KL;
 PI WPI; 2001-041427/05.
 DR
 XX Genomic library for identifying diagnostic and therapeutic
 PT compositions, and for identifying virulence factors, regulatory
 PT elements and drug targets, comprises Moraxella catarrhalis nucleic
 PT acids -
 PS Claim 1; Page 223-235; 545pp; English.
 XX
 XX The present invention relates to a Moraxella catarrhalis genomic library
 CC comprising of a combination of 41 nucleic acid molecules (see
 CC AAF28514-AAF28554). The library has a number of uses described in the
 CC specification e.g. is useful for identifying diagnostic and therapeutic
 CC compositions. M. catarrhalis (Branhamella catarrhalis) is a large
 CC aerobic, gram-negative diplococcus, normally found among the bacterial
 CC flora of human upper airways. M. catarrhalis is known to cause acute,
 CC localised infections such as otitis media, sinusitis and bronchopulmonary
 CC infection and life-threatening, systemic diseases including endocarditis
 CC and meningitis.
 XX
 SQ Sequence 48328 BP; 14211 A; 9322 C; 10802 G; 13992 T; 1 other;
 Alignment Scores:
 Pred. No.: 2.09e-114 Length: 48328
 Score: 1670.00 Matches: 322
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 22 Gaps: 0
 US-10-018-706-2 (1-322) x AAF28540 (1-48328)
 QY 1 MetThrValThrIleAlaIleAsnSerGlnAsnGlnValProIleValArgLeuGlyLeu 20
 Db 40850 ATGACTGTGAGATGACATCAATTCACAAATCAAAACCATCAACGATGGGCTTG 40909
 QY 21 IlePheGlyValIleThrThrCysIleLeuAlaGlyCysAlaSerValProThrTyrAsn 40
 Db 40910 ATTTTGGTGTGATCACCACTTCATTGGCAGATGTGCCAGTAACCAACCTATAT 40969
 QY 41 SerThrSerGlySerGlySerHisArgThrSerGlySerGlyLeuAlaIleGlySer 60
 Db 40970 AGTACCTCAGGTTGGGCGAGCATGCTTCAAGTTCAAGGTTGGCAATAGTTCA 41029
 QY 61 GlnValIleThrAspSerGlnGlyValProAsnArgTyrGlnValValGlnGlyAspThr 80
 Db 41030 CAGGTTATCAGGACGACGACGAGCGGTACCAATCCCTATCAGTGAACGAGGCGCATCT 41089
 QY 81 ValSerValIleAlaGlnArgTyrGlyLeuAsnTyrPArgGluIleGlyHisIleAsnAsn 100
 Db 41090 GTCCAGTACGATTGCTCAGCGTTATGATTAATTCGCGAGATTGGACACATTATAT 41149
 QY 101 LeuAsnSerSerTyrThrIleTyrThrGlyIleProLeuThrLeuTyrSerGlyAspLeu 120
 Db 41150 CTAATATGACAGTTATATGATTTATACAGTCAATGCGTCACTTATGTGATGATCTC 41209
 QY 121 LysValAlaGlyGluArgSerIleSerSerGlyValAsnThrAlaHisThrProSerProVal 140
 Db 41210 AAGGTGCGTGAAGCGATGATCAGCTGTGTGATTAACGCTCAACACCTTCGCTGTG 41269
 QY 141 AlaValGlnSerSerArgProProValGlnGlnHisProAlaValGlnLysProThrPro 160
 Db 41270 GCGGTTGATGACAGACGACCACTACGACGATCTGCGGTACAAAAACCCACGCGCA 41329
 QY 161 ProValValValValValValSerProThrProThrProProValValGlnGlnProAlaPro 180

Db 41330 CCTGTTGTTGTGTAACCAACCAACCAACCACTCCGCTGTGTTACAGACGACACCA 41389
 QY ValAlaProProValThrGluAlaProPheAlaThrGlySerSerGlyValMetGlnPhe 200
 Db 41390 GTTGCCCAACAGTACAGACAGACACATTTTCCACGCGGTAGCTCAGGGGTGATGCAATTT 41449
 QY 201 ArgTyrProValGlyAlaThrAsnProValValAlaArgArgPheGlyThrAlaThrValAla 220
 Db 41450 CGCTATCTGTTGGTGGACCAATCCAGTGGTGGACCTTTGTGACGCGACAGTGGCC 41509
 QY 221 GlySerThrValThrSerSerGlyMetTyrPheSerGlyValGAspGlyAspLeuIleAsn 240
 Db 41510 GGCTCACTGTATACAGTAAATGCGATGTGTCTGTGACAGATGGCGGATTTAATTAC 41569
 QY 241 AlaSerAsnAlaGlyThrValIleGlnAlaAspHisAsnMetAspGlyAlaSerIleVal 260
 Db 41570 GCCAGTAATGACGAGGACGATATTCACCTGATCAATATGAGAGCGGGCCAGTATGTG 41629
 QY 261 IleGlnHisThrAsnGlyPheValSerSerTyrIleHisIleLysAspAlaGlnValLys 280
 Db 41630 ATTACGATACCAATGATTTGTTTCAAGCTATATTCATATTAAGACGCTCAAGTTAA 41689
 QY 281 ThrGlyAspThrValArgThrGlyGlnArgIleAlaSerMetLysAsnGlnProSerGly 300
 Db 41690 ACAGCGATACGAGCGGTACCGGTACGCTATTTGCAACAGAAATCAGCCAAACCGCT 41749
 QY 301 AlaAlaLeuPheGlnPheArgIleSerArgAsnGlyValTyrValAspProLeuThrVal 320
 Db 41750 GCGGCACTATTGAAATTTAGAAATTTCTAGAAATGCGGTGATGTGATTCATGACAGTA 41809
 QY 321 LeuLys 322
 Db 41810 CTTAAA 41815
 RESULT 3
 ID AAF30047 standard; DNA; 966 BP.
 XX AAF30047;
 AC AAF30047;
 DT 23-APR-2001 (first entry)
 XX
 DE Moraxella catarrhalis DNA encoding BASB110 protein.
 KW BASB110; infection; otitis media; pneumonia; gene therapy;
 KW diagnosis; antibacterial; antimicrobial; genetic immunisation;
 KW vaccine; ss.
 OS Moraxella catarrhalis.
 XX
 XX
 PH Key location/Qualifiers
 FT conflict replace(310,A)
 FT /*tag a
 FT /note= "PCR product differs from gene sequence
 FT of AAF30046 at this base"
 XX
 XX WO200100838-A1.
 XX
 PD 04-JAN-2001.
 XX
 PF 23-JUN-2000; 2000WO-EP05854.
 XX
 PR 25-JUN-1999; 99GB-0015031.
 XX
 PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
 PI Thonard J;
 XX
 DR WPI; 2001-112459/12.
 DR P-P8DB; AAB20107.
 XX
 XX Novel BASB110 polypeptides of Moraxella catarrhalis, useful as a
 PT vaccine for treating Moraxella catarrhalis infections -

XX Claim 10; Page 83; 88pp; English.

XX
 CC The present sequence is that of DNA encoding the BASB110 protein
 CC (see AAB20107) of Moraxella catarrhalis Mc2931 (ATCC 43617), a
 CC causative agent of otitis media in children and pneumonia in adults.
 CC The DNA was obtained by PCR amplification of Mc2931 plasmid DNA
 CC (see also AAF30048-49). The sequence differs from that of the
 CC BASB110 gene coding region sequence (see AAF30046) by having G at
 CC position 310 to Ser in the translated sequence. The invention
 CC provides BASB110 polypeptides, and polynucleotides encoding them,
 CC as well as expression vectors, host cells and methods for producing
 CC BASB110 polypeptide using recombinant methods. Also claimed is a
 CC vaccine composition comprising a BASB110 polypeptide, an immunogenic
 CC fragment of a BASB110 polypeptide, or a polypeptide having at least
 CC 85% amino acid sequence identity to BASB110, or comprising a
 CC polynucleotide encoding such a polypeptide. A claimed method of
 CC diagnosing a Moraxella infection involves identifying a BASB110
 CC polypeptide or antibody. A claimed therapeutic composition useful
 CC in treating humans with M. catarrhalis infection comprises at least
 CC 1 antibody directed against a BASB110 polypeptide. BASB110
 CC polynucleotides also have utility in diagnosis of the stage and
 CC type of infection, and also for therapeutic or prophylactic
 CC purposes, in particular genetic immunisation.

XX Sequence 966 BP; 263 A; 214 C; 242 G; 247 T; 0 other;

Alignment Scores:

Pred. No.: 3,71e-116 Length: 966
 Score: 1666.00 Matches: 321
 Percent Similarity: 99.69% Conservative: 0
 Best Local Similarity: 99.69% Mismatches: 1
 Query Match: 99.76% Indels: 0
 DB: 22 Gaps: 0

US-10-018-706-2 (1-322) x AAF30047 (1-966)

QY 1 MetThrValThrIleAlaIleAsnSerGlnIleValProAsnArgTyrGlnValLeu 20
 DB 1 ATGACTGTGACGATGACATCATTCACAAATGCAATGCAATGCAATGCAATGCAATG 60
 QY 21 IlePheGlyValIleThrThrCysIleLeuAlaGlyCysAlaSerIleProThrTyrAsn 40
 DB 61 ATTTTGTGTGATCACCACCTTGATTTGCGAGATGTCCAGTAAGCCAACTTAAT 120
 QY 41 SerThrSerGlySerGlySerHisArgThrSerGlySerGlyLeuAlaIleGlySer 60
 DB 121 AGTACCTCAGGTCGGGACGATCATCTTCAGTTCAGGTGGTTGGCAATAGGTTCA 180
 QY 61 GlnValIleThrAspSerGlnIleValProAsnArgTyrGlnValLeuAspThr 80
 DB 181 CAGGTTATCAGCAGCAGTCAAGGCGGTACCAATGCAATGCAATGCAATGCAATG 240
 QY 81 ValSerIleIleAlaGlnArgTyrGlyLeuAsnTrpArgIleGlyHisIleAsnAsn 100
 DB 241 GTGAGTAAGATTGCTCAGCGTTATGATTAATGGCGTGAAGATTGACACATTAAAT 300
 QY 101 LeuAsnSerSerTyrThrIleTyrThrGlyGlnTrpLeuThrLeuTrpSerGlyAspLeu 120
 DB 301 CTAAATAGCGGTATATAGATTATACAGTCATAGCTGCTTATATGTCAGGTGATCTC 360
 QY 121 LysValAlaArgIleArgSerIleSerSerGlyValAlaAsnThrAlaHisTrpSerProVal 140
 DB 361 AAGGTGCGTAGCGATGATACAGCTGTGCTGTGTGATACAGCTCACACACCTTGCGCTGTG 420
 QY 141 AlaValGlnSerSerArgProProValGlnGlnHisProAlaValGlnIleProThrPro 160
 DB 421 GCGGTTCATGTCAGACACACACAGTACAGAGATCTCCCGTACAAAAACCCACGCCA 480
 QY 161 ProValValValValValValValValValValValValValValValValValValVal 180
 DB 481 CCTGT 540

QY 181 ValAlaProProValThrGlnAlaProPheAlaThrGlySerSerGlyValMetGlnPhe 200
 DB 541 GTTGCCCCACAGTGCACAGAACACCATTTGCCAGGGAGCTCAGGGGTGTGCAATTT 600
 QY 201 ArgTyrProValAlaValAlaThrAsnProValValAlaArgAspGlyThrAlaThrValAla 220
 DB 601 CGGTATCTGTTGGTGGCGACCATCATCAGTGGTTGCACCGCTTGTGTACGGCGACAGTGCC 660
 QY 221 GlySerThrValThrSerAsnGlyMetTrpPheSerGlyArgAspGlyAspLeuIleAsn 240
 DB 661 GGTCAACAGTTATCCAGTAATGCAATGTGGTTTCTGTGACGAGATGGGCAATTAACTAAC 720
 QY 241 AlaSerAsnAlaGlyThrValIleGlnAlaAspHisAsnMetAspGlyValAspThrIleVal 260
 DB 721 GCGAGTAATCAGGACAGCAGTCTTCAAGCTGATCATATATGACGCGGCGATTTGTG 780
 QY 261 IleGlnHisThrAsnGlyPheValSerSerTyrIleHisIleLysAspAlaGlnValLys 280
 DB 781 ATTCAGCATACCAATGAGATTGTTCAGACTATATTCATTTAGAGAGCTCAAGTTAA 840
 QY 281 ThrGlyAspThrValArgThrGlyGlnArgIleAlaSerMetLysAsnGlnProSerGly 300
 DB 841 ACAGGGGATACCGTGTGTCACCGTCAAGGTATTCAGAGCATGAAAAATCAGCAAGCGGT 900
 QY 301 AlaAlaLeuPheGlnPheArgIleSerArgAsnGlyValIleValAspProLeuThrVal 320
 DB 901 GCGGCACTATTGAAATTGTAGAAATTCGTAGAAATGCGGTGTATGTGATTCATTCAGAGTA 960
 QY 321 LeuLys 322
 DB 961 CTTAA 966

RESULT 4
 AAT42063
 ID AAT42063 standard; DNA; 1830121 BP.
 AC AAT42063;
 XX 14-SEP-1999 (first entry)
 DE Haemophilus influenzae complete genome sequence.
 KW Genome; bacterium; Haemophilus influenzae; computer readable medium;
 KW expression modulating fragment; regulation; gene expression; vector;
 KW organism; open reading frame; ORF; ds.
 XX Haemophilus influenzae.
 OS
 XX MO9633276-A1.
 PN
 XX 24-OCT-1996.
 PD
 XX 22-APR-1996; 96WO-US05320.
 PF
 XX 07-JUN-1995; 95US-0487429.
 PR 21-APR-1995; 95US-0426787.
 PR 07-JUN-1995; 95US-0476102.
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA (UYXO) UNIV JOHNS HOPKINS.
 XX
 XX Adams MD, Fleischmann RD, Smith HO, Venter JC, White O;
 XX WFI; 1996-485782/48.
 XX
 XX Haemophilus influenzae Rd genome recorded on computer readable
 PT medium - useful for identifying commercially important nucleic acid
 PT fragments by homology searching
 XX
 XX Claim 1; Page 77.2-77.1091; 1291pp; English.
 PS
 CC This sequence represents the complete genome sequence of the bacterium

Percent Similarity: 44.09%
 Best Local Similarity: 24.80%
 Query Match: 12.51%
 DB: 21
 Gaps: 6

US-10-018-706-2 (1-322) x AA53349 (1-2179)

```

OY 73 TyrGlnValIleGlnGlyAspThrValSerLysIleAlaGlnArgTyrGlyLeuAsnTrp 92.
DB 1229 TCAAGATTCGCAAGGCGATTCATGCTTCTATTGCTTATATTCAGCGATGATATA 1288
OY 93 ArgGluIleGlyHisIleAsnLeuAsnSerSerTyrThrIleTyrThrGlyGlnTrp 112
DB 1289 AAGAAATTCGCAACTATATATGTCGTGAGCCCATCATCTGATTTGGACAGATA 1348
OY 113 LeuThrLeuTrpSerGly-----AspLeuLysValArgGluArgSerIleSerSer 129
DB 1349 TTGAAATTCGCAATATATATTCCTCCGATAGCAATATGATCCAAACAGACAAATTAATGAA 1408
OY 130 GlyValAsnThrAlaHisThrProSerProValAlaValGlnSerSerArgProProVal 149
DB 1409 TCAGAGGTGACCAAAATACACTCAATAG--ACATGGAATGCTAAATAAACCAACAAT 1465
OY 150 GlnGlnHisProAlaValGlnLysProThrProProProValValValLysLysProThr 169
DB 1466 GAACAAATGAAACCCGTCGCTACACCAACATTCATTCACATTCATCAATTAATTA----- 1519
OY 170 ProThrProProValValGlnGlnProAlaProValAlaProProValThrGluAlaPro 189
DB 1520 ---ACACCTCCGACCACTCAATATAGCTTGATTTGGCCA----- 1558
OY 180 PheAlaThrGlySerSerGlyValMetGlnPheArgTyrProValGlyAlaThrAsnPro 209
DB 1559 -----ACAAATGGAATA 1570
OY 210 ValValArgArgPheGlyThrAlaThrValAlaGlySerThrValThrSerAsnGlyMet 229
DB 1571 ATTATTCAGAGATTTTCAGTCTCATAGAGGC-----ATTAAGGATTT 1615
OY 220 TrpPheSerGlyArgAspGlyAspLeuIleAsnAlaSerAsnAlaGlyThrValIleGln 249
DB 1616 GATATTACCGGTCCTCGGACACAGCTGTAATGACAGCGCTGGACGAGGTGTATAT 1675
OY 250 AlaAspHisAsnMetAspGlyAlaSer-----IleValIleGlnHisThrAsnGlyPhe 267
DB 1676 GCCGAGACGCTTACGTGATATGTAATTTAATTTATTTAAACATATGACAGTTAT 1735
OY 268 ValSerSerTyrIleHisIleLysAspAlaGlnValLysThrGlyAspThrValArgThr 287
DB 1736 TTAAGTGTATGACCATATGAAAGTATCTCGTCAAGATTCAGCAAGAAATTAAAGCG 1795
OY 288 GlyGlnArgIleAlaSerMetLysAsnGlnProSerGlyAlaAlaLeuPheGlnPheArg 307
DB 1796 GGTCAACAAATTCCTAAATTCGGAAGTCTGGAACAAACAAATCAATCAATTCATTGAA 1855
OY 308 IleSerArgAsnGlyValTyrValAspProLeuThrValLeu 321
DB 1856 ATTCGTTATTAAGTCATCATGATGATTCATCAATGAGATATTATA 1897

```

RESULT 6
 AA53075
 ID AA53075 standard; DNA; 1248 BP.
 XX AA53075;
 XX
 XX 21-MAR-2000 (first entry)
 XX
 DE Neisseria meningitidis ORF 025 partial DNA sequence SEQ ID NO:111-1.
 XX
 XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
 KM antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
 XX antibacterial; gene therapy; ds.

```

OS Neisseria meningitidis.
XX
XX MO9957280-A2.
XX
XX 11-NOV-1999.
XX
XX 30-APR-1999; 99MO-US09346.
XX
XX 01-MAY-1998; 98US-0083758.
XX 31-JUL-1998; 98US-0094869.
XX 02-SEP-1998; 98US-0098994.
XX 02-SEP-1998; 98US-0099062.
XX 09-OCT-1998; 98US-0103749.
XX 09-OCT-1998; 98US-0103794.
XX 09-OCT-1998; 98US-0103796.
XX 25-FEB-1999; 99US-0121528.
XX
XX (CHIR ) CHIRON CORP.
XX (GENO-) INST GENOMIC RES.
XX
XX Fraser C, Galeotti C, Grandi G, Hickey E, Maignani V, Mora M,
XX Petersen J, Pizsa M, Rappuoli R, Ratti G, Scalato E, Scarselli M,
XX Tettelin H, Venter JC,
XX WPI; 2000-062150/05.
XX P-PSDB; AAY74313.
XX
XX Novel Neisserial polypeptides predicted to be useful antigens for
XX vaccines and diagnostics -
XX
XX Claim 7; Page 206-207; 1453pp; English.
XX
XX AA53015 to AA54536, AA54577 to AA54615, and AAY74253 to AAY75941
XX represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
XX and polypeptides. AA54537 to AA54576 and AA54616 to AA54713 represent
XX PCR primers used in the exemplification of the present invention. The
XX polypeptides, the polynucleotides, antibodies and compositions of
XX the invention can be used as vaccines, as diagnostic reagents, and as
XX immunologic compositions. The polypeptides can be used in the
XX manufacture of medicaments for treating or preventing infection due to
XX Neisserial bacteria (e.g. meningitis and septicemia), to detect the
XX presence of Neisseria bacteria, or to raise antibodies. They may also
XX be used to screen for agonists or antagonists, which may themselves
XX have use as antibacterial agents. The polynucleotides of the invention
XX may also be used in gene therapy protocols.
XX
XX Sequence 1248 BP; 277 A; 387 C; 331 G; 251 T; 2 other;
XX
XX Alignment Scores:
XX Pred. No.: 2,786-06 Length: 1248
XX Score: 203.00 Matches: 71
XX Percent Similarity: 43.46% Conservative: 42
XX Best Local Similarity: 27.31% Mismatches: 97
XX Query Match: 12.16% Indels: 50
XX DB: 21 Gaps: 11
XX
XX US-10-018-706-2 (1-322) x AA53075 (1-1248)
OY 73 TyrGlnValIleGlnGlyAspThrValSerLysIleAlaGlnArgTyrGlyLeuAsnTrp 92
DB 553 CATACATTCGTGCGGCGGACGCGGTACACATTTCCAAAGGCTACCATATCTGTCAA 612
OY 93 ArgGluIleGlyHisIleAsnLeuAsnSerSerTyrThrIleTyrThrGlyGlnTrp 112
DB 613 GACGATTTCCGTCGCGGAAACGCGATGACCGCAAT--ACGTGAGCATCGGTGAG-- 666
OY 113 LeuThrLeuTrpSerGlyAspLeuLysValArgGluArgSerIleSerSerGlyValAsn 132
DB 667 -----ATTGTTAAAGTCAAAACCGCGAGATATGCCCA----- 699
OY 133 ThrAlaHisThrProSerProValAlaValGlnSerSerArgProProValGlnGlnHis 152
DB 700 -----CCGAAAGCCGACGCCGTAAAA--AGCAGGCCCGCGGTACCG----- 738

```

Qy	153	ProAlaValGlnLysProThrProProValValValLysLysProThrProThrPro	172
Db	739	GCTGCGGTGCAAAACCCCTGTG-----AAACCCGCGCGCAACCG	777
Qy	173	ProValValGlnGlnProAlaProValAlaPro-----	183
Db	778	CCTGTGCAGTCCGCGCGCAACCTCCGCGCGCGGAAATAAAGCGGTTCCGCG	837
Qy	184	ProValThrGlnAlaProPheAla-----ThrGlySerSerGlyValMetGln	199
Db	838	CCCCGCCCGCAATCTCTGCGCTTCGCTCCGCGCACGCGCTTCGTCGCGCATTTGT	897
Qy	200	PheArgTyrProValGlyAlaThrAsnProValValArgAspPheGlyThrAlaThrVal	219
Db	898	TGGCAGCGCTCG-----ACGCAAGGTAAAGTGTGTCGCAATTTCGCGCGC-----	942
Qy	220	AlaGlySerThrValThrSerAsnGlyMetTrpPheSerGlyArgAspGlyAspPheIle	239
Db	943	-----AACAAACAAGGTGTGCAATTATGCAGGAAATAGCGGACAGCCGCT	987
Qy	240	AsnAlaSerAsnAlaGlyThrValIleGlnAlaAspHisAsnMetAspGlyAlaSer---	258
Db	988	TTGGGGGGCGCTGACGGCAAAAGTGTATATGCAGGTTCCGTTGTAGGGGATACCGCAAT	1047
Qy	259	---IleValIleGlnHisThrAsnGlyPheValSerSerTyrIleHisIleLysAspAla	277
Db	1048	TTGGTCATCATTCACGATATATTTCTTCTTCGACCGCATACGGGACACCAACAAATTG	1107
Qy	278	GlnValLysThrGlyAspThrValArgThrGlyGlnArgIleAlaSerMetLysAsnGln	297
Db	1108	CTGGTCGGCGAAGGCCACAGGTCAAAAGCGGGCAGCAGTCCGTTTANTGGGCAATAC	1167
Qy	298	ProSerGlyAlaAlaLeuPheGlnPheArgIleSerAlaGlnGlyValTyrValAspPro	317
Db	1168	GAGGCTTCTTGAAGACGACGCTTCATTTCGAGGTGCGGCAAAAACGGCAACCGGTTAATCCG	1227
RESULT 7			
AAZ53073			
ID	AAZ53073	standard; DNA; 1014 BP.	
XX	AAZ53073;		
AC			
XX			
DT	21-MAR-2000	(first entry)	
XX			
DE	Neisseria gonorrhoeae ORF 025 partial DNA sequence SEQ ID NO:117.		
XX			
KW	Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine; antigenic; diagnosis; immunogenic; infection; meningitis; septicemia; antibacterial; gene therapy; ds.		
XX			
OS	Neisseria gonorrhoeae.		
XX			
PN	W09957280-A2.		
XX			
PD	11-NOV-1999.		
XX			
PE	30-APR-1999; 99WO-US09346.		
XX			
PR	01-MAY-1998; 98US-0083758.		
PR	31-JUL-1998; 98US-0094869.		
PR	02-SEP-1998; 98US-0098994.		
PR	02-SEP-1998; 98US-0099062.		
PR	09-OCT-1998; 98US-0103749.		
PR	09-OCT-1998; 98US-0103794.		
PR	09-OCT-1998; 98US-0103796.		
PR	25-FEB-1999; 99US-0121528.		
XX			
PA	(CHIR) CHIRON CORP.		
XX	(GENO-) INST GENOMIC RES.		
PI	Fraser C, Galeotti C, Grandi G, Hickey E, Massignani V, Mora M, Petersen J, Piza M, Rappuoli R, Ratti G, Scarlato E, Scarselli M,		

PI Tectelin H, Venter JC;
XX WPI; 2000-062150/05.
DR P-PsDB; AA74311.

Novel Neisserial polypeptides predicted to be useful antigens for
vaccines and diagnostics -

Claim 7, Page 205-206; 1453pp; English.

AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
represent novel *Neisseria meningitidis* and *N. gonorrhoeae* polynucleotides
and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
PCR primers used in the exemplification of the present invention. The
polypeptides, the polynucleotides, antibodies and compositions of
the invention can be used as vaccines, as diagnostic reagents, and as
immunogenetic compositions. The polypeptides can be used in the
manufacture of medicaments for treating or preventing infection due to
Neisserial bacteria (e.g. meningitis and septicemia), to detect the
presence of *Neisseria bacteria*, or to raise antibodies. They may also
be used to screen for agonists or antagonists, which may themselves
have use as antibacterial agents. The polynucleotides of the invention
may also be used in gene therapy protocols.

SQ Sequence 1014 BP; 209 A; 318 C; 271 G; 216 T; 0 other;

Alignment Scores:

Pred. No.:	3,63e-06	Length:	1014
Score:	200.00	Matches:	66
Percent Similarity:	44.05%	Conservative:	45
Best Local Similarity:	26.19%	Mismatches:	109
Query Match:	11.98%	Indels:	32
DB:	21	Gaps:	9

US-10-018-706-2 (1-322) x AAZ53073 (1-1014)

73 TyrgInValIyVGInGIYAAsPThYrAlSerlySIIeAlGlnArGYrGLyeAunTP 92
::: :
313 CATACTATTGGCCGGCGGCAGACAGGTGTACAACATTTTCCAAACGCCATCACCATTCCTCA 372
93 ArgGnIIleGIHIAIEAsmAsmEunsenSerTYrThIIeyTrhGrGIunTP 112
::: :
373 GACGATTCCTCGTGCGTGAACCGCATGACCACAA--ACGTGACATCGGTCAg--- 426
113 LeuthLeutPrSeRGIYAAspleuYSyVAIArgGIunGrSerIIeserGeRGIYValasn 132
427 -----ATTGTTAAAGTCAAAACCGCAGCATATGCCACCGAATAACC 466
133 ThrAlAHsthrProSerProVaIAIalValnSerSerArXprProVaIGInIHIs 152
::: :
469 GCAGCCGTAGAAGACAGGCCCGCCGTACCGCGTCCGCCAACACCTGTGA----- 522
153 ProIAValAIgInlyPrOthrProrProVaIVAlvalIyalySerhPrOthrPro 172
::: :
523 CCCGCCGCGCAACCGCCC-----GTTCAGTCCGCGCGCAACCTGCGCG 567
173 ProVaIVaIGInI-----ProIAProVaIAIAProProVaIThGrGuaIa--- 188
568 CCCGCTGGCGAAATAAAGCGGTTCCGCCGCCGCCGCCGCCGCCGAAATCTCTGCCGCT 627
189 ---ProphealathrGIyserserGIYalMetGlnphearGYrPrOVAIGIyalatnr 207
628 TCgcCTTCcCGGcAcCGcGTTCGcgTcGcgGcGcATGTTTggcAcGcGTcGc-----ACCAA 681
208 AsnProVaIVaIARgArPhelyIThrAlatnrVALalaglyserThrVal]hrsraSn 227
682 GGTAAGTGTGGCCGATTTCCGCGCGCGC-----AACAAg 717
228 GlymetTrphneserGIYArgAspSIYAspleuIIeasAlaserASnaIagIyThrVal 247
718 GTGTGCATATTGccGcGcATncCGGcAcACccCGTTTggcGcGcGcGTACGCGAAAgTs 777
248 IleginalAspHIAsmMetaspGIYalaser-----IeValIIegInHsthrAsn 265

AA81766;
 04-DEC-2000 (first entry)
 N. meningitidis partial DNA sequence gnm_313 SEQ ID NO:313.
 Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;
 antigen; vaccine; diagnosis; infection; antibacterial; identification;
 Meningococcus B; Memb; ds.
 Neisseria meningitidis:
 WO200022430-A2.
 20-APR-2000.
 08-OCT-1999; 99WO-US23573.
 09-OCT-1998; 98US-0103794.
 30-APR-1999; 99US-0132068.
 (CHIR) CHIRON CORP.
 Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC,
 Maignani V, Galeotti C, Mora M, Ratti G, Scarcelli M, Scarlato V,
 Rappunoli R, Pizzo M;
 WPI; 2000-318079/27.
 Isolated nucleotide sequences of Neisseria meningitidis which can be
 used in the diagnosis and treatment of N. meningitidis infection and
 other Neisserial infections, for example, N.gonorrhoea -
 Claim 7; Page 1588-1590; 1760pp; English.
 The present invention describes methods of obtaining immunogenic
 proteins from Neisseria genomic sequences. AA81453 to AA82414
 represent specifically claimed Neisseria meningitidis genomic DNA
 sequences; AA81260 to AA81303 and AB825620 to AB825663 represent
 Neisseria DNA sequences and their corresponding proteins; AA81254 to
 AA81259 and AA81304 to AA81321 represent PCR primers used in the
 isolation of Neisseria meningitidis DNA sequences; and AA81322 to
 AA81452 represent Neisseria meningitidis Memb polynucleotide ORF
 sequences, which are all used in the exemplification of the present
 invention. The nucleic acid sequences, protein sequences, and antibodies
 against them, can be used in the manufacture of a composition. The
 composition can be used as a medicament (or in the manufacture of a
 medicament) for treating, preventing or diagnosing infection due to
 Neisserial bacteria. For example, some of the identified proteins could
 be components of vaccines against Meningococcus B; against all serotypes
 and/or against all pathogenic Neisseriae. Identification of sequences
 from the bacterium will also facilitate production of biological probes,
 particularly organism-specific probes. Attempts to make efficacious
 Meningococcus B vaccines have failed mainly due to antigen tolerance.
 Multivalent vaccines have also been tried but none have successfully
 overcome antigenic variability. The provision of further, complete
 sequences may provide an opportunity to identify secreted or surface
 exposed proteins that may be presumed targets for the immune system and
 which are not antigenically variable or at least more conserved than
 other more variable regions.

Percent Similarity: 43.80%
 Best Local Similarity: 23.55%
 Query Match: 9.88%
 DB: 14
 Gaps: 6

US-10-018-706-2 (1-322) x ABQ51084 (1-1885)

QY TyGlnValIysGlnGlyAspThrValSerIleAlaGlnArgIleGlyLeuAsnTrp 92
 1229 TACAAGTACGCAAGGAGATGATGTTCTTATGTTGTTATTTGATTCAGGATGATATA 1288
 QY ArgGluIleGlyIleAsnAsnLeuAsnSerSerIleIleIleIleIleIleIleIle 112
 1289 AAGAAATGGCAACATAATATATATGTCAGCCATATCATCTGATATGGACAAGTA 1348
 QY LeuThrIleuTrpSerGly-----AspLeuIysValAlaGlnArgSerIleSerSer 129
 1349 TTGAAATTCGCAATTAATATTCGCAATGCAATATGATGACCAACAGACAAATGAA 1408
 QY GlyValAsnThrAlaHisThrProSerProValAlaValGlnSerSerArgProProVal 149
 1409 TCAGAGTGACAAATATACGTCATGAG--ACATGATGCTTATTAACCAACAAT 1465
 QY GlnGlnHisProAlaValGlnIleProThrProProValValValValIleIleProThr 169
 1466 GAACAAATGAACCCGTTGTCACACCAACATTCACCAATGCCAATCAATTAAT 1519
 QY ProThrProProValValGlnGlnProAlaProValAlaProProValThiGlnAlaPro 189
 1520 ---ACACCTCAGCCACCTCAATATATAGCTTGATTTGGCCA----- 1558
 QY PheAlaThrIleSerSerGlyValMetGlnPheArgIleProValGlyAlaThrAsnPro 209
 1559 -----ACAAATGGAAA 1570
 QY ValValArgArgPheGlyThrAlaThrValAlaGlySerThrValThrSerAsnGlyMet 229
 1571 ATTAATTCAGATTTTCACAGTGTGATGAGGC-----AATTAAGTATT 1615
 QY TrpPheSerGlyArgAspGlyAspLeuIleAsnAlaSerAsnAlaGlyThr--ValIleG 249
 1616 GATATTAAGTCTTATGCAATATGCAAGCTGTTAATGCAAGCTGATGAGCAGCTTAT 1675
 QY LnaIleAspHisAsnMetAspGlyValAser-----IleValIleGlnHisThrAsnGlyP 267
 1676 ATGCCGAGAGACGCTTACGTCGATATGTAATTTAATTTAATTAACATATACAGCTT 1735
 QY heValSerSerIleIleHisIleLeuAspAlaGlnValIleIleIleIleIleIleIleIle 287
 1736 ATTTAAGTCTTATGCAATATGCAAGCTGTTAATGCAAGCTGATGAGCAGCTTAAAG 1795
 QY hrgIleGlnArgIleAsnMetIleAsnGlnProSerGlyValAlaLeuPheGlnPhe 307
 1796 CGGATCAACAATTTGTAATATGGAAGTCTGGAACAACAATCAATCAATCAATTTTA 1855
 QY 307 rg 307
 Db 1856 AA 1857
 RESULT 14
 ABQ67196 standard; DNA, 684707 BP.
 XX ABQ67196;
 AC ABQ67196;
 DT 29-AUG-2002 (first entry)
 XX Listeria innocua contig DNA sequence #9.
 DE Listeria innocua contig DNA sequence #9.
 XX Antibacterial; Listeria; food contamination; mutational analysis;
 KM infection; de.
 XX Listeria innocua.

XX WO200228891-A2.
 PN 11-APR-2002.
 XX 04-OCT-2001; 2001WO-FR03061.
 XX 04-OCT-2000; 2000FR-0012697.
 XX 04-OCT-2000; 2000FR-0012697.
 PA (INSP) INST PASTEUR.
 PA (CNRS) CNRS CENT NAT RECH SCI.
 PI Kunst F, Glaser P;
 DR WPI; 2002-332479/37.
 XX New genomic sequences from Listeria species, useful for detection,
 PT treatment and prevention of infection, also related polypeptides,
 PT antibodies and modulators
 PS Claim 5; SEQ ID 9; 180bp; French.
 CC The present invention relates to nucleic acid sequences
 CC (ABQ67188-ABQ71212) from Listeria sp. The sequences are useful as probes
 CC and primers for identification and/or detection of Listeria (e.g. as
 CC contaminants in foods, or mutational analysis) and for analysis of
 CC gene expression. Proteins encoded by the nucleic acid sequences can be
 CC used to screen for compounds that modulate gene expression, replication
 CC and pathogenicity of Listeria (potential therapeutic agents), also for
 CC treating infections by Listeria, and are useful as immunogens in
 CC anti-Listeria vaccines.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic format
 CC directly from WIPO at ftp.wipo.int/pub/published_pcf_sequences.
 XX SQ Sequence 684707 BP; 213423 A; 126375 C; 133438 G; 211468 T; 3 other;
 Alignment Scores:
 Pred. No.: 12.2 Length: 684707
 Score: 158.50 Matches: 77
 Percent Similarity: 36.09% Conservative: 41
 Best Local Similarity: 23.55% Mismatches: 127
 Query Match: 9.49% Indels: 82
 DB: 24 Gaps: 13
 US-10-018-706-2 (1-322) x ABQ67196 (1-684707)
 QY 27 ThrCysIleLeuAlaGlyCysAlaSerIleProThrTyraSerThrSerGlySerGly 46
 Db 615017 AGTTGTATC---CGCAATGTTTAAAGCGACCAAAATATATATGCAATCAAAAAGACGA 615073
 QY 47 SerHisArgThrSerGlySer-----GlyGly 55
 Db 615074 TCAAAATTAACATAAAGTACAGAAATATGTCAGAAAAAATTAAGAAACTTAAAGT 615133
 QY 56 -LeuAlaIleGlySerGlnValIleThrAspSerGlnGlyValProAsnArgTyrglnVa 75
 Db 615134 GCTAGCAGTTGAGCTTACAGTA-----TCAAGAAACAACATGGAAAGCCAAACA 615184
 QY 75 IlyGlnGlnIleAspThrValSerIleAlaGlnArgIleGlyLeuAsnTrpArgGluI 95
 Db 615185 AGAAAGAAATGACTTATGATGCTGTCGCAATTAATAAAGACTTAAG----- 615233
 QY 95 eGlyHisIleAsnAsnLeuAsnSerSerIleIleIleIleIleIleIleIle 115
 Db 615234 -----AAAAGTGAACAACAACCTTTTAACATATGA----- 615263
 QY 115 uTrpSerGlyAspLeuIysValAlaGlnArgSerIleSerSerGlyValAsnThrAlaHis 135
 Db 615264 -----CAAGTGCTTTTCAAGACCAACAAGAAACTGCTTCAATATTTGCTGGTGA 615319
 QY 135 sThrProSerProValAlaValGlnSerSerArgProProValGlnGlnHisProAlaVa 155

```

Db      615320 AGCAAAACAGAGCTGCAATTAAGCGGCTGAGAAAGAAACGATGCAAGAGCAGCGGC 615379
Oy      155 1GlnlysProThrProProValValValValLysLysProThrProThrProProValVal 175
Db      615380 TAAAAATGCGCACT-----GTGCGCAGC 615400
Oy      175 1GlnGlnProAlaProValAlaProProValThrGluAlaProPhaAlaThrGlySerSe 195
Db      615401 GCAGCAACCGAGCTGTGAACATCTTCACAGAGAACTGGCGCAGACGACGATAGTTC 615460
Oy      195 rGlyValMetGlnPheArgTyrProValGlyAla----- 206
Db      615461 TGGTGGCGGCGCACTTTATCAAGCGACATCAAGAAATTTAACTGCACTGATTTAGTACGC 615520
Oy      207 -ThrsnProValValArgArgPhe-----GlyThrAlaThrValAlaGlySe 222
Db      615521 TACTAACCGATTACTGGAATAATATCAATCTCAATAAGGTCAGATATTGCGGCGGAGG 615580
Oy      222 rThrValThrSerAsnGlyMetTrpPheSerGlyArgAspGlyAspLeuIleAsnAlaSe 242
Db      615581 TACGATTACA-----GTATCCGCTGC 615601
Oy      242 rAsnAlaGlyThrValIleGlnAlaAspHisAsnMetAspGlyAlaSer----- 258
Db      615602 GGCATCTGTAGAGTGTGATTTTTCAGGATTGGCGCACAGGTAGCGGCTTGGAGGTTA 615661
Oy      259 -----1LevalIleGlnHisThrAsnGlyPheValSerSerTyrIleHisIlely 275
Db      615662 TGGTACGTCGTGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 615721
Oy      275 sAspAlaGlnValLys-----ThrGlyAspThrValArgThrGlyGlnArgIleAlaSe 293
Db      615722 TCCCGGTAGCTGAAAGTAGTAAGTGTCAACAGATACCAAGGACCAACCATTTGGTAT 615781
Oy      293 rMet-----LysAsnGlnProSerGlyValAlaLeuPheGluPheArgIleSerArgAs 311
Db      615782 TATGGGATCCACCGGTCATCAACGACGACACATCTT---CATTTGAGATTTCACAAAAA 615838
Oy      311 nGlyValTyrValAspPro 317
Db      615839 TGGTATTCACGTTGATCCA 615857

RESULT 15
ABA03041/C
ID      ABA03041 standard; DNA; 2944528 BP.
XX      ABA03041;
AC      05-FEB-2002 (first entry)
DT      05-FEB-2002 (first entry)
XX      Listeria monocytogenes EGD-e genome sequence.
XX      Anticbacterial gene therapy; vaccine; biosynthesis; biodegradation;
XX      vitamin B12; bacterial infection; disease; ds.
XX      Listeria monocytogenes.
XX      WO200177335-A2.
XX      18-OCT-2001.
XX      11-APR-2001; 2001WO-FR01118.
XX      11-APR-2000; 2000FR-0004629.
XX      (INSP) INST PASTEUR.
XX      Buchrieser C, Frangeul L, Couve E, Rusniok C, Esahi H, Denoux P,
PI      Dussuguet O, Chetoui F, Nedjari H, Glaeser P, Kunst F, Cossart P,
PI      Daniels J, Goebel W, Krefel J, Kuhn M, Ng E, Vazquez-Boland JA,
PI      Dominguez-Bernal G, Garrido-Garcia P, Tiberet-Martinez A, Amend A,
PI      Chakraborty T, Dommann E, Hain T, Berche F, Charbit A, Durant L,
PI      Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;

```

```

PI      Maduenio E, De Pablos B, Wehlant J, Kaerst U, Entian K, Hauf J;
PI      Rose M, Vose H;
DR      WPI: 2002-010914/01.
XX      Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
PT      and prevention of Listeria and related bacterial infections, and
PT      related polypeptides
XX      Claim 1; SEQ ID No 1; 192bp; French.
XX      The present sequence is the genome sequence of Listeria monocytogenes
CC      EGD-e. This sequence and fragments of this sequence are useful for
CC      selecting probes and primers for detecting genes in L. monocytogenes and
CC      related organisms, and to study genetic polymorphisms and other genomes.
CC      Proteins (AB047297-AB050149) expressed from the present sequence are
CC      useful for raising specific antibodies, identification of L.
CC      monocytogenes and related organisms, and for biosynthesis and
CC      biodegradation, especially biosynthesis of vitamin B12. This sequence and
CC      proteins encoded by it are also useful for selecting compounds that
CC      regulate gene expression and cell replication and modulate L.
CC      monocytogenes-related diseases. In addition, this sequence and proteins
CC      encoded by it are useful in pharmaceutical and vaccine compositions for
CC      the treatment or prevention of infections by L. monocytogenes and related
CC      organisms.
CC      Note: The sequence data for this patent did not form part of the printed
CC      specification, but was obtained in electronic format directly from WIPO
CC      at ftp.wipo.int/pub/published_pcr_sequences.
SQ      Sequence 2944528 BP; 914202 A; 563301 C; 555061 G; 911964 T; 0 other;

Alignment Scores:
Pred. No.: 71.2 Length: 2944528
Score: 158.50 Matches: 66
Percent Similarity: 37.22% Conservative: 33
Best Local Similarity: 24.81% Mismatches: 100
Query Match: 9.49% Indels: 67
DB: Gaps: 11

US-10-018-706-2 (1-322) x ABA03041 (1-2944528)
Oy      76 LysGlnGlyAspThrValSerLysIleAlaGlnArgTyrGlyLeuAsnTrpArgGluIle 95
Db      2581153 GAAAAAATGATTTAGTTATGCGCTTTCAGCAACAAAAAGACTTAAC----- 2581106
Oy      96 GlyHisIleAsnAsnLeuAsnSerSerTyrThrIleTyrThrGlyGlnTrpLeuThrLeu 115
Db      2581105 -----AAGAGTGAACAAACCTTTTACCTAGTGAA----- 2581076
Oy      116 TrpSerGlyAspLeuLysValArgGluArgSerIleSerSerGlyValAlaLeuThrAlaHis 135
Db      2581075 ---CAAGTGCCTTACTGATGTAAGAAAAAAGACTTCTTAATGCTGCGTGAAGAAA 2581019
Oy      136 ThrProSerProValAlaValAlaGlnSerSerArgProProValGlnGlnHisProAlaVal 155
Db      2581018 GCAAAACAGAGCTGCAATTAAGTCAAGTCCGGAAGAAAAAGAGATGCAAGAGTCCGCGCA 2580959
Oy      156 GlnLysProThrProProValValValValLysLysProThrProThrProProValVal 175
Db      2580958 GCATCAAGCGCTTAAGAGTGGCGCTGTGTCAA----- 2580926
Oy      176 GlnGlnProAlaProValAlaProProValThrGluAlaProPhaAlaThrGlySerSer 195
Db      2580925 ---CAACCAAGTTCGTCATCAATAAGACGACGAGAA-----ACAGTAAAGTTCC 2580881
Oy      196 GlyValMetGlnPheArgTyrProValGlyAla----- 206
Db      2580880 GGGTGAAGCGCAATTTTAAACAGGCTTGTGATTTTAACCTTGCATTTAGTGAACGA 2580821
Oy      207 ThrAsnProValValArgArgPhe-----GlyThrAlaThrValAlaGlySer 222
Db      2580820 ACAATTCCTCCGTAGTGCAGAAATCGAATCCATTAAGGTCAGATATTGCTGTGTGTGTG 2580761

```

QY 223 ThrValThrSerAsnGlyMetTrpPheSerGlyArgAspGlyAspLeuIleAsnAlaSer 242
 DB 2580760 ACAGTTACT-----GTTTCAGCTGCA 2580740
 QY 243 AsnAlaGlyThrValIleGlnAlaAspHisAsnMetAspGlyAlaSer----- 258
 DB 2580739 GCATCCGGAACAGTTGATATTTCTGGGTTGGTCATCGAGTATGGAATTTGGTGTAT 2580680
 QY 259 -----IleValIleGlnHisThrAsnGlyPheValSerSerTyrlleHisIleLeu 275
 DB 2580679 GCGTATGTAGTAAAAATTGACCAAGCGGTTCCAGACACATATATGACATATGGCT 2580620
 QY 276 AspAlaGlnValIle-----ThrGlyAspThrValArgThrGlyGlnArgIleAlaSer 293
 DB 2580619 GCAAGCAGTTTGAAGTATACCGGTCAACAGATTTCACAGGCCCAACCAATTGGTATC 2580560
 QY 294 Met-----LysAsnGlnProSerGlyAlaAlaLeuPheGluPheArgIleSerArgAsn 311
 DB 2580559 ATGGGATCAACAGGTCATCGACAGCAACATCTA---CATTTGAATACATATAAAC 2580503
 QY 312 GlyValTyArgValAspPro 317
 DB 2580502 GGTATTCAGTCAATCCA 2580485
 RESULT 16
 ABO69245/c
 ID ABO69245 standard; DNA; 3011208 BP.
 XX
 AC ABO69245;
 XX
 DT 29-AUG-2002 (first entry)
 XX
 DE Listeria innocua DNA sequence #684.
 XX
 KM Antibacterial; Listeria; food contamination; mutational analysis;
 KM infection; ds.
 XX
 OS Listeria innocua.
 XX
 PN WO200228891-A2.
 XX
 PD 11-APR-2002.
 XX
 PF 04-OCT-2001; 2001WO-FR03061.
 XX
 PR 04-OCT-2000; 2000FR-0012697.
 XX
 PA (INSP) INST PASTEUR.
 PA (CNRS) CNRS CENT NAT RECH SCI.
 XX
 PI Kunst F, Glaser P;
 XX
 DR WPI; 2002-332479/37.
 XX
 PT New genomic sequences from Listeria species, useful for detection,
 PT treatment and prevention of infection, also related polypeptides,
 PT antibodies and modulators -
 XX
 PS Claim 5; SEQ ID 2058; 180pp; French.
 XX
 CC The present invention relates to nucleic acid sequences
 CC (ABO67188-ABO67192) from Listeria sp. The sequences are useful as probes
 CC and primers for identification and/or detection of Listeria (e.g. as
 CC contaminants in foods, or mutational analysis) and for analysis of
 CC gene expression. Proteins encoded by the nucleic acid sequences can be
 CC used to screen for compounds that modulate gene expression, replication
 CC and pathogenicity of Listeria (potential therapeutic agents), also for
 CC treating infections by Listeria, and are useful as immunogens in
 CC anti-Listeria vaccines.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic format
 CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 3011208 BP; 941651 A; 568176 C; 559189 G; 942192 T; 0 other;
 Alignment Scores:
 Pred. No.: 73.1 Length: 3011208
 Score: 158.50 Matches: 77
 Percent Similarity: 36.09% Conservative: 41
 Best Local Similarity: 23.55% Mismatches: 127
 Query Match: 9.49% Indels: 82
 DB: 24 Gaps: 13
 US-10-018-706-2 (1-322) x ABO69245 (1-3011208)
 QY 27 ThrCysIleLeuAlaGlyCysAlaSerLysProThrTyraSerSerThrSerGlySerGly 46
 DB 2665796 AGTTGTATC---CGCAATGTTTAAAGCGAACCAAAATATATGCAAGATCAAAAAGCGA 2665740
 QY 47 SerHisArgThrSerGlySer-----GlyGly 55
 DB 2665739 TCAAAATTAACCTAAAGTAGCAGAAATATGTTCAAAAAAAATTAAGAACTTAAGAGT 2665680
 QY 56 -LeuAlaIleGlySerGlnValIleThrAspSerGlnGlyValProAsnArgTyArgIleVal 75
 DB 2665679 GCTAGCAGTTGAGCTTAAAGTA-----TCAAGAAACAACATGGAAAGCCAAACA 2665629
 QY 75 LlyGlnGlyAspThrValSerLysIleAlaGlnArgTyArgIlePheAsnTrpArgGluI 95
 DB 2665628 AGAAAGAAATGACTTATGATGGCTCTTGCAAAATAAAGACTTAAG----- 2665580
 QY 95 eGlyHisIleAsnAsnLeuAsnSerSerTyrlleTyrlleTyrlleTyrlleTyrlle 115
 DB 2665579 -----AAAGTGAAACAACACTTTAATCTAATGAA----- 2665550
 QY 115 uTrpSerGlyAspLeuLysValArgGluArgSerIleSerSerGlyValAsnThrAlaHi 135
 DB 2665549 ----CAAGTGCTCTTTCACAGCAAGAAAGAAACCTTCTTAATATGCTGTGTAATA 2665494
 QY 135 sThrProSerProValAlaValGlnSerSerArgProProValGlnGlnHisProAlaVal 155
 DB 2665493 AGCAAAACAAAGAACGCTCAATTAAAGCGCTGAAGAAACGCTGCAAGAGACAGCGCG 2665434
 QY 155 lGlnLysProThrProProValValValLysLysProThrProThrProProValVal 175
 DB 2665433 TAAAAATGCCACT-----GTGCGACG 2665413
 QY 175 lGlnGlnProAlaProValAlaProProValThrGlnAlaProPheAlaThrGlySerSe 195
 DB 2665412 GCAGCAACCAAGCTCTGTATCATCTTCAGCAGAACTGGCGCACTGACACAGTAAGTTC 2665353
 QY 195 rGlyValMetGlnPheArgTyrlleProValGlyAla-----GTATCCGCTGC 206
 DB 2665352 TGGTGGCGGCGAGTTTATCAAGCCAGCATCAGAAATGTTAACATCTGATTTAGTGACGG 2665293
 QY 207 -ThrAsnProValValArgArgPhe-----GlyThrAlaThrValAlaGlySe 222
 DB 2665292 TACTAACCCAGTTACTGAAATAATATGATCATTAAGGTCACAGATATTTGGCGGCGAGG 2665233
 QY 222 rThrValThrSerAsnGlyMetTrpPheSerGlyArgAspGlyAspLeuIleAsnAlaSe 242
 DB 2665232 TACGATTACA-----GTATCCGCTGC 2665212
 QY 242 rAsnAlaGlyThrValIleGlnAlaAspHisAsnMetAspGlyAlaSer----- 258
 DB 2665211 GGCATCTGTAAGAGCTGATATTTTCAGATTGGCGCAACAGTACGGCTTTGAGAGTTA 2665152
 QY 259 -----IleValIleGlnHisThrAsnGlyPheValSerSerTyrlleHisIleLeu 275
 DB 2665151 TGGTACGCTGTAAGAAATATGATCATGCTTATGCTTCAACAACATTATATGCTATATGCG 2665092
 QY 275 sAspAlaGlnValIle-----ThrGlyAspThrValArgThrGlyGlnArgIleAlaSe 293
 DB 2665091 TGGCGGTAGCTTGAAGTATGATGCTCAACAGATATCAAGAGAACCAACCAATTGGTAT 2665032
 QY 293 rMet-----LysAsnGlnProSerGlyAlaAlaLeuPheGluPheArgIleSerArgAs 311

```

Db 265031 TATGGGATCCACCGGTCAATCAACCGCAACAATCTT---CATTTGAGATTCACAAAAA 2664975
QY 311 ngiVal1TyValAspPro 317
Db 2664974 TGGTATTCAGTTGATCCA 2664956

RESULT 17
ABA90521/c
ID ABA90521 standard; DNA; 2365589 BP.
XX
AC ABA90521;
XX
DT 16-MAY-2002 (first entry)
XX
DE Genomic sequence of Lactococcus lactis IL1403.
XX
KM Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese; ds.
XX
OS Lactococcus lactis IL1403.
XX
PN FR2807446-A1.
XX
PD 12-OCT-2001.
XX
PF 11-APR-2000; 2000FR-0004630.
XX
PR 11-APR-2000; 2000FR-0004630.
XX
PA (INRG ) INRA INST NAT RECH AGRONOMIQUE.
XX
PI Bolotine A, Sorokine A, Renault P, Ehrlich SD;
XX
DR WPI; 2002-043418/06.
XX
PT New nucleotide sequence useful in the identification or Lactococcus
XX
PS lactis and related species -
XX
XX
XX Claim 1; SEQ ID 1; 2504bp; French.
XX
CC The present invention is related to a Lactococcus lactis nucleotide
CC sequence (ABA90521) and related proteins (AB53300-AB55521). The
CC nucleic acid sequence is useful in the detection and/or amplification of
CC related species. The proteins of the invention are useful for the
CC biosynthesis or biodegradation of a composition of interest. The
CC invention helps research in lactic bacteria, particularly useful in the
CC production of yogurt and cheese.
CC Note: The sequence data for this patent is based on equivalent patent
CC WO20017734 (published 18-OCT-2001) which is available in electronic
CC format directly from WIPO at tcp.wipo.int/pub/published_pat_sequences.
XX
SQ Sequence 2365589 BP; 765914 A; 415261 C; 420487 G; 763927 T; 0 other;

Alignment Scores:
Pred. No.: 398 Length: 2365589
Score: 147.00 Matches: 77
Percent Similarity: 38.22% Conservative: 43
Best Local Similarity: 24.52% Mismatches: 112
Query Match: 8.80% Indels: 83
DB: 24 Gaps: 12

US-10-018-706-2 (1-322) x ABA90521 (1-2365589)
QY 3 ValThr1leAla1leAsnSerGlnAsnGlnLysProIleLys-----16
Db 269567 ATTAGCTAAAGCTTGGCGAGCAACGCGCACTTCAAGCTGCAACCGCAGCCTTAA 269508
QY 17 -----ArgLeuGly1leu1lephGlyVal1leThrTrpGly1leuAlaGly 32
Db 269507 CAGGAAATATGCAACTGACCTATATAT-GGTGCTTCCCGAATCGAATATATTTCT--- 269452
QY 33 CysAlaSerLysProThrTyAsnSerThr-----SerGlySerGlySerHisArgThr 50

```

```

Db 269451 -----CAATATATTTAGCTGCTTTTACAGGCTCTTCTGCTGACT 269407
QY 51 SerGlySerGlyValLeuAla1leGlySerGlnVal1leThrAspSerGlyValPro 70
Db 269406 TCTATTTCCGGTGT---TCAACAGCTTCAAAATCAATCAATTAATCAATCAAGCTCA 269350
QY 71 AsnArgTyTrpGlnVal1leGly1leAspThrValSerLys1leAlaGlnArgTyGlyLeu 90
Db 269349 ACCACTTATACAGATTAATCTGGGGAATCACTTTGGGAATTTGCACAAATATGGAATT 269290
QY 91 AsnTrpArgGlu1leGly1leAsnAsnLeuAsnSerSerTyThr1leTyThrGly 110
Db 269289 AGTGTGCTCAAAATTCAAAGCGCAACAAATCTTAAAGTACA---GTCACTTATATTGGG 269233
QY 111 GlnTrpLeuThrLeuTrpSerGlyAspLeuLysValArgGluArgSerLysSerSerGly 130
Db 269232 CAAAAGCTTGTATGACAACTTCAAGTCTTCTGCTAATCAAAATAGTCAACTTCTTCA 269173
QY 131 ValAsnThrAla1leThrProSerProValAlaValGlnSerSerArgProProValGln 150
Db 269172 GGAATTTCTGCGGAGACTACAAACGCTACTACT----- 269140
QY 151 GlnHisProAlaValGlnLysProThrProProValValValLysLysProThrPro 170
Db 269139 -----TCGGTCACTCCT 269128
QY 171 ThrProProValValGlnGlnProAlaProValAlaProProValThr-----186
Db 269127 GCCAAACGAGCTTACAGACAGACAGATTAAGTTAAATCTGATAGCTTTGGGACTC 269068
QY 187 GluAlaProPheAlaThrGlySerSerGlyValMetGlnPheArgTyTrpProValGlyAla 206
Db 269067 TCTGTCAATTAATAAAGACAGATTCCTCACTCAAGTGGATCAATTTGAATTCGAT 269008
QY 207 ThrAsnProValValArgTyPheGlyThrAlaThrValAlaGlySerThrValThrSer 226
Db 269007 ACAATTTTCATGGACAAATCTGATTTGTTTCAATCTCCGATCTTCAAGTCTTCA 268948
QY 227 AsnGlyMetTrpPheSerGlyArgAspGlyAspLeu1leAsnAlaSerAlaGlyThr 246
Db 268947 ACAAGT-----TCAGGCTCAGGCTCTTACG 268924
QY 247 Val1leGlnAlaAspHisAsnMetAspGlyAlaSer1leVal1leGln1leThrAsnGly 266
Db 268923 -----AGTTCAACTTCTTAATCTTCTGACGCTTCA-----268894
QY 267 PheValSerSerTyTrp1leHis1leLysAspAlaGlnValLysThrGlyAspThrVal---285
Db 268893 -----AATACCTCTATCCAT-----AAGTTGTAAAGAGATACGCTTTGG 268852
QY 286 -----ArgThrGlyGlnArg1leAlaSerMetLys 295
Db 268851 GGACTTTCACAAAATCTGTAGCCCAATGCTTCAATTAAAG 268810

RESULT 18
AAH67706
ID AAH67706 standard; DNA; 615 BP.
XX
XX AAH67706;
XX
DT 26-SEP-2001 (first entry)
XX
DE C glutamicum coding sequence fragment SEQ ID NO: 2741.
XX
KM Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
XX
OS organic acid synthesis; ds.
XX
XX Corynebacterium glutamicum.
XX
XX EPI108790-A2.
XX
XX 20-JUN-2001.

```

XX 18-DEC-2000; 2000EP-0127688.
 XX 16-DEC-1999; 99JP-0377484.
 PR 07-APR-2000; 2000JP-0159162.
 PR 03-AUG-2000; 2000JP-0280988.
 XX (KYOW) KYOWA HAKKO KOGYO KK.
 PA Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H,
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
 DR WPI; 2001-376931/40.
 XX P-PSDB; AAG92487.
 PT Novel polynucleotides derived from Coryneform bacteria, for identifying
 PT mutation point of a gene, measuring expression of a gene, analysing
 PT expression profile or pattern of a gene and identifying homologous gene
 PT
 PS Claim 8; SEQ ID NO: 2741; 246bp + Sequence Listing; English.
 XX
 XX The present invention provides a number of nucleotide and protein
 CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
 CC are useful for identifying the mutation point of a gene derived from a
 CC mutant of coryneform bacterium, measuring expression amount and
 CC analysing the expression profile or expression pattern of a gene derived
 CC from Coryneform bacterium, and identifying a homologue of a gene derived
 CC from coryneform bacterium. Coryneform bacteria are useful for producing
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
 CC particularly L-lysine. The present sequence is a nucleic acid described
 CC in the exemplification of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC European Patent Office.
 XX
 SQ Sequence 615 BP; 148 A; 229 C; 139 G; 99 T; 0 other;
 Alignment Scores:
 Pred. No.: 0.0204 Length: 615
 Score: 146.50 Matches: 64
 Percent Similarity: 44.44% Conservative: 20
 Best Local Similarity: 33.86% Mismatches: 67
 Query Match: 8.77% Indels: 39
 DB: 22 Gaps: 10
 US-10-018-706-2 (1-322) x AAH67706 (1-615)
 QY 137 ProSerProValAlaValGlnSerSerArgProProValGlnGlnHisProAlaValGln 156
 DB 107 CCCCAACCGCTACCG-----CCACCATGGGACCGACCATCG-----CGC 145
 QY 157 LysProThr-----ProProValAlaValValValValValValValValValValVal 172
 DB 146 AAACCGGATGAGGAGGCGCTTCCCTCCGCGCTG-----CCGACACCGCTCCGACG 193
 QY 173 ProValValGln---GlnProAlaProValAlaProProValThnGlnAlaProPheAla 191
 DB 194 AAGCTGAGAGATGCGACACCGCCGCTC-GCACACCGCCGACCGGACGCGCCGACCA 252
 QY 192 ThrGlySerSerGlyValMetGlnPheArgTyrProValGlyAlaThrAspProValVal 211
 DB 253 AACGGGACCC-----TTCACCTCAGAGATTGGA 279
 QY 212 ArgArgPheGlyThrAlaThrValAlaGlySerThrValThrSerArgMetTyrPhe 231
 DB 280 CCAAGTTGGGAAAC-----TTCACCAACGCGATCGACATC 315
 QY 232 SerGlyArgAspGlyAspLeuIleAsnAlaSerAsnAlaGlyThrValIleGlnAlaAsp 251
 DB 316 GCAACATCATTCGCGACCCCAATCATGCGCTCATGCGCGACATGATGAGTCTGGC 375
 QY 252 HisAsnMetAsp---GlyAlaSerIleValIleGlnHisThrAsnGlyPheValSerSer 270

DB 376 CCAAGATCCGGCTATGACAGTGGATCCGATCCAGACGACGACGATCATCTCCATC 435
 QY 271 TyrIleHisIleLeuAspAlaGlnValLysThrLysPheThrValArgThrGlyGlnArg 290
 DB 436 TACGACACATGAGAAATCCTTACGCTCTCCGTCGCGCAACGCGTGGACGACGAGGAA 495
 QY 291 IleAlaSerMetLysAsnGlnPro---SerGlyAlaAlaLeuPheGluPheArgIleSer 309
 DB 496 ATCCAGAGATGGGACACCAAGATTCTCCACCGGCTCCACCTCCACTTGAATTCAC 555
 QY 310 ArgAsnGlyValTyr---ValAspPro 317
 DB 556 CCAAGCGCGCTCACCCCGATCGACCA 582
 RESULT 19
 AAH68532
 ID AAH68532 standard; DNA; 349980 BP.
 XX
 AC AAH68532;
 XX
 DT 26-SEP-2001 (first entry)
 XX
 DE C glutamicum coding sequence fragment SEQ ID NO: 7067.
 XX
 KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
 KW organic acid synthesis; ds.
 XX
 OS Corynebacterium glutamicum.
 XX
 PN EP108790-A2.
 XX
 PD 20-JUN-2001.
 XX
 XX 18-DEC-2000; 2000EP-0127688.
 PF 16-DEC-1999; 99JP-0377484.
 PR 07-APR-2000; 2000JP-0159162.
 PR 03-AUG-2000; 2000JP-0280988.
 XX
 PA (KYOW) KYOWA HAKKO KOGYO KK.
 PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H,
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
 DR WPI; 2001-376931/40.
 XX
 XX Novel polynucleotides derived from Coryneform bacteria, for identifying
 PT mutation point of a gene, measuring expression of a gene, analysing
 PT expression profile or pattern of a gene and identifying homologous gene
 PT
 PS Disclosure; SEQ ID NO: 7067; 246bp + Sequence Listing; English.
 XX
 XX The present invention provides a number of nucleotide and protein
 CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
 CC are useful for identifying the mutation point of a gene derived from a
 CC mutant of coryneform bacterium, measuring expression amount and
 CC analysing the expression profile or expression pattern of a gene derived
 CC from Coryneform bacterium, and identifying a homologue of a gene derived
 CC from coryneform bacterium. Coryneform bacteria are useful for producing
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
 CC particularly L-lysine. The present sequence is a nucleic acid described
 CC in the exemplification of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC European Patent Office.
 XX
 SQ Sequence 349980 BP; 80900 A; 98397 C; 92139 G; 78544 T; 0 other;
 Alignment Scores:
 Pred. No.: 43.3 Length: 349980
 Score: 146.50 Matches: 64

Percent Similarity: 44.44%
 Best Local Similarity: 33.86%
 Query Match: 8.77%
 DB: 22
 Gaps: 10

US-10-018-706-2 (1-322) x AAF59421 (1-349980)

DB 137 ProserProValAlaValGlnSerSerThrProProValGlnGlnHisProAlaValGln 156
 DB 249656 CCGCAACCGCTACCG-----CCACCAAGGACCGCAATCG-----CGC 249694
 QY 157 LysProThr-----ProProValAlaValValValValValValValValValValVal 172
 DB 249695 AAACCGACTACGAGCGCTTTCCTCCGCGCTG-----CCGACACCGCTCGCAG 249742
 QY 173 ProValValGln--GlnProAlaProValAlaProProValThrGlnAlaProPheAla 191
 DB 249743 AAGCTGAGAGTGCACCAACCGCGCTC--GCACCAAGCGCGCACTGAGCGCGCCAGCA 249801
 QY 192 ThrGlySerSerGlyValMetGlnPheArgTyrProValGlyAlaThrAspProValVal 211
 DB 249802 AACGGCAC-----TTCCACTCAGGATTCGGA 249828
 QY 212 ArgArgPheGlyThrAlaThrValAlaGlySerThrValThrSerArgGlyMetTyrPhe 231
 DB 249829 CCAGCTGGGGGAC-----TTCCACCAAGCGCATGCATC 249864
 QY 232 SerGlyArgAspGlyAspLeuLeuAlaSerAspAlaGlyThrValIleGlnAlaAsp 251
 DB 249865 GCAAATCAATCGGCACACCCCAATCTACGCTCATGCGCGCATCTGCATCAGCTCGC 249924
 QY 252 HisAspMetAsp---GlyAlaSerIleValIleGlnHisThrArgGlyPheValSerSer 270
 DB 249925 CCAGATCCGGCTATGACAGTGCATCGCATCGACGACGACGACGATCATCTCATC 249984
 QY 271 TyrIleHisIleLysAspAlaGlnValLysThrGlyAspThrValArgThrGlyGlnArg 290
 DB 249985 TACGGACACATGAAATACCTCTACGCTCTCCGCGCAACCGCTGCAGCAGCGCGCAAGA 250044
 QY 291 IleAspSerMetLysAspGlnPro--SerGlyAlaAlaLeuPheGlnPheArgIleSer 309
 DB 250045 ATCCACAGAAATGGCGAGCGCAAGATTCCTCCACCGCTCCCATCTCGAATCCAC 250104
 QY 310 ArgAspGlyValTyr---ValAspPro 317
 DB 250105 CCAGACCGCGCTCACCCCGATGCACCA 250131

RESULT 20
 AAF59421/c
 ID AAF59421 standard; DNA; 949 BP.
 XX AAF59421;
 AC AAF59421;
 DT 02-MAY-2001 (first entry)
 XX
 DE Actinobacillus actinomycetemcomitans clone nucleotide sequence #8.
 XX
 KM Actinobacillus actinomycetemcomitans; microbial; infection; vaccine;
 KM Identification; localised juvenile periodontitis; antibacterial;
 KM antiinflammatory; de.
 XX
 OS Actinobacillus actinomycetemcomitans.
 XX
 MO200111081-A2.
 XX
 PD 15-FEB-2001.
 XX
 PF 04-AUG-2000; 2000WO-US21340.
 XX
 PR 06-AUG-1999; 99US-0147551.
 XX
 PA (IVIG-) IVIGENE CORP.

PI Progulsk-Fox A, Handfield M, Brady LJ, Hillman JD;
 DR WPI; 2001-202779/20.
 XX

PT Identifying microbial polynucleotides, useful for vaccine design,
 PT diagnostic and antibodytherapy, comprises isolating clones of a
 PT microbe's expression library, reactive with antibodies against microbe
 PT proteins produced during in vivo growth.

XX Example 3; Page 63; 68pp; English.

XX The present invention describes a method (M1) for identifying a
 CC polynucleotide (I) of a microbe (M) that is expressed in vivo. (M1)
 CC comprises: (a) absorbing antibodies (Ab) against antigens that are
 CC expressed by (M) in vivo and in vitro with cells or cellular extracts
 CC of (M) that have been grown in vitro; (b) isolating unadsorbed Abs; and
 CC (c) probing an expression library of (M)'s DNA/RNA with unadsorbed Abs,
 CC where (I) that is expressed in vivo is identified. The method can be
 CC used for identifying antigens expressed during an actual microbial
 CC infection. The identified polynucleotides are useful for vaccine design,
 CC diagnostic and antibodytherapy, in particular for the diagnosis and
 CC therapy of Actinobacillus actinomycetemcomitans infection, which is the
 CC etiologic agent for localised juvenile periodontitis. The present
 CC sequence represents an Actinobacillus actinomycetemcomitans clone
 CC nucleotide sequence, which is used in an example from the present
 CC invention.

SQ Sequence 949 BP; 170 A; 236 C; 281 G; 261 T; 1 other;

Alignment Scores:

Pred. No.: 0.0447 Length: 949
 Score: 145.00 Matches: 57
 Percent Similarity: 37.04% Conservative: 23
 Best Local Similarity: 26.39% Mismatches: 62
 Query Match: 8.68% Indels: 74
 DB: 22 Gaps: 9

US-10-018-706-2 (1-322) x AAF59421 (1-949)

QY 73 TyrGlnValLysGlnGlyAspThrValSerLysIleAlaGlnArgTyrGlyLeuAsnTyr 92
 DB 537 TATACCGTACGCAAGGTGCACCATGTCTTATCGCTTACATTTCTGTTTGGATGTG 478
 QY 93 ArgGluIleGlyHisIleAspAsnLeuAsnSerSerTyrThrIleTyrThrGlyGlnTyr 112
 DB 477 AAAGATTGGCGCGCTGAATATATGTCGCAACCTTACAGCTTAAGGTGCAGCAAAAC 418
 QY 113 LeuThrLeuTyrSerGlyAspLeuLysValArgGluArgSerIleSerSerGlyValAsn 132
 DB 417 TTA-----AAGTGAGCGCGCGC----- 400
 QY 133 ThrAlaHisThrProSerProValAlaValGlnSerSerArgProProValGlnGlnHis 152
 DB 399 -----ACAAAGGTTCAGGTGCAGAAAGAAAGCTGCACCCGCAATATGCA--- 358
 QY 153 ProAlaValGlnLysProThrProProValValValValValValValValValValVal 172
 DB 357 -----CCAGCGCTCACCA 343
 QY 173 ProValValGlnGlnProAlaProValAlaProProValThrGlnAlaProPheAlaThr 192
 DB 342 CCGGCTACGCAAGGTGTGACACCG-----GTAACTTACACCCCGCGCGCAAC 292
 QY 193 GlySerSerGlyValMetGlnPheArgTyrProValGlyAlaThrAspProValValArg 212
 DB 291 GGCACA-----CAATATGTGTTCCAGCGCAGCATTAACCGCGCGCGTAAGCC 244
 QY 213 ArgPheGlyThrAla-----ThrValAlaGlySerThr--- 223
 DB 243 GCGGTAGTACTGTGCACCGCAACCAATACGCGATTAACCAAGCGCGCGCACCGCACCG 184
 QY 224 ValThrSerAspGlyMetTyr-----PheSer 232

Db 183 GCAGTATCAACGTGGCATGGCGTGGCCGACCAACGGCAATGATTCAGATTCTCT 124
Qy 233 G1yAryAspGly-----AspLeuIleAsn 240
Db 123 AACGCCGAGCGCGGTACAAAGCATGCATCACTAGTGTCTCCGACAGCAAGCATCAT 64
Qy 241 AlaseerAsnAlaGlyThrValIleGlnAlaAspHisAsnMetAspGly 256
Db 63 GCCGCCGCCGACAGGTGAGTGTGTTACCGGGGTAAAGCCTTACGCGGT 16

RESULT 21
AAx20500/c
ID AAx20500 standard; DNA; 14063 BP.
XX
AC AAx20500;
XX
XX 05-MAY-1999 (first entry)
XX
XX Polynucleotide sequence from the genome of Treponema pallidum.
XX DE Treponema pallidum infection; syphilis; Borrelia infection; animal;
XX KW enzyme production; ds.
XX
XX Treponema pallidum.
XX OS
XX MO9859034-A2.
XX PN
XX 30-DEC-1998.
XX PD
XX 23-JUN-1998; 98WO-US13041.
XX PF
XX 24-JUN-1997; 97US-0050667.
XX PR
XX (HUMA-) HUMAN GENOME SCI INC.
XX PA
XX
XX Fraser CM;
XX PI
XX WPI; 1999-081273/07.
XX DR
XX
XX New isolated Treponema pallidum nucleic acids - used to develop
XX PT products for the detection, diagnosis, characterisation, prevention
XX and therapy of T. pallidum infections, particularly syphilis
XX
XX
XX Claim 1; Page 178-186; 1150pp; English.
XX
XX AAx20500-21243 represent polynucleotide sequences from the genome of
XX CC Treponema pallidum. The sequences can be used for detection,
XX CC diagnosis, characterisation, prevention and therapy for T. pallidum
XX CC infections, particularly syphilis. They can also be used for detecting
XX CC diseases related to Borrelia infections in animals, and for the
XX CC production of biosynthetic products such as enzymes.
XX
XX
XX Sequence 14063 BP; 3225 A; 4086 C; 3507 G; 3225 T; 20 other;
XX SQ

Alignment Scores:
Pred. No.: 2.75 Length: 14063
Score: 140.00 Matches: 82
Percent Similarity: 38.62% Conservative: 52
Best Local Similarity: 23.63% Mismatches: 141
Query Match: 8.38% Indels: 73
DB: 20 Gaps: 14

US-10-018-706-2 (1-322) x AAx20500 (1-14063)

Qy 23 G1yVallIleThrThrCysIleLeuAlaGlyCysAlaSerIleProThrTyrAsnSerThr 42
Db 8443 GGAACATTGACACCTGCTCCATCAGCTCATGAGCA-----GGAACACAG 8399
Qy 43 SerGlySerGly-----SerHisArgThrSerGlySerGly 54
Db 8398 TCAGCGCTGCGGAGTACGTGTCTTCTTCTGCTCCGTCGACGTATCGTGCAGAAAGATGCGGA 8339
Qy 55 GlyLeuAlaIleGlySerGlnValIleThrAspSerGlnGlyValProAsnArgTyrGln 74

Db 8338 GCTGCACTGACTACTACACAGTGA-----CGA 8312
Qy 75 ValIleSerGlnGlyAspThrValSerIleAlaGlnArgTyrGlyLeuAsnThrArgGln 94
Db 8311 ATGCGTGAAGGTGATGTGTGTGAGAAAGATTCGACGCTATGACATGACAGAGATCA 8252
Qy 95 IleGlyHisIleAsnAsnLeuAsnSerSerTyrThrIleTyrGlnGlyIleThr 114
Db 8251 ATCATTAATTGAATTAATTTCGACACACAGG-GCGCTCCAGTTGGACACTACTAAAA 8193
Qy 115 LeuTrpSerGlyAsp-----LeuIleValArgGlnArgSerIleSerSerGlyVal 131
Db 8192 ATCCCTCAGTGAAGCAGCATTTTATATACCGTAAACGCGCATACGTTTCCGTATTA 8133
Qy 132 AsnThrAlaHisThrProSerProValAlaValGlnSerSerArgProProValGlnGln 151
Db 8132 GCGCGCGCGCAT-----CAGATCTCTTAGAACATGTCGCTGCTTAAC 8088
Qy 152 HisProAlaValGlnIleProThrProProValVal-----ValValIleLysPro----- 168
Db 8087 ACGCGCTTCTTCAAGAGATCACTCTTCACTGCTGACGCTGTTCTCCGTCTAT 8028
Qy 169 -----ThrProThrPro----- 172
Db 8027 AATTGAGTGCACGAGAGTCTCGGTGCTTTTCGTTTCTTTCGCAACAGTGAAG 7968
Qy 173 -----ProValValGln-----GlnProAlaProValAlaProProVal 185
Db 7967 GAAATACGTGCTTGTGACGACGATGACGATGACGATGACGATGACGATGACGATGAC 7920
Qy 186 ThrGlnAlaProPheAlaThrGlySerSerGlyValMetGln----- 199
Db 7919 CTGTTTTCGCGGTGACATTAAAGCGACGCGCGGTGACGAAATTAATGTCGACTTG 7860
Qy 200 PheArgTyrProValGlyAlaThrAsnProValValArgArgPheGly-----ThrAla 217
Db 7859 TTTCGTGCTCTCTTCGACGATGATCATGTTCTTCGCGGTATGATGATGATGATGATGAT 7800
Qy 218 ThrValAlaGlySerThrValThrSerAsnGlyMetThrPheSerGlyArgAspGlyAsp 237
Db 7799 CCGTTTACCGGTGACAGAGTTTTCACATGTTGTGATGATGATGATGATGATGATGATGAT 7740
Qy 238 LeuIleAsnAlaSerAsnAlaGlyThrValIleGlnAlaAspHisAsn-----MetAspGly 256
Db 7739 CCAAGTATATTCGGGCTCGGACGAGGATTTGACCAACGTTGATACAGTGCATGATGAT 7680
Qy 257 AlaseerIleValIleGlnHisThrAsnGlyPheValSerSerTyrIleHisIleLysAsp 276
Db 7679 AATTACCTGATTGTGGGACACACGCGGGGTATACAGACCCGTATNGGCACTGACAGAG 7620
Qy 277 AlaGlnValIleThrGlyAspThrValArgThrGlyGlnArgIleAlaSerMet---Lys 295
Db 7619 GTGTGTGTTTCAGACGATCCGCGCTACACCGCCGACCAAAATCGATTTATTTGGAAAA 7560
Qy 296 AsnGlnProSerGlyAlaAlaLeuPheGlnPheArgIleSerArgAsnGlyValIleVal 315
Db 7559 ACAGAGCGACACAGGACACACACCTGATTTTCAATCTATTAAGATGAGTCCGCGGATA 7500
Qy 316 AspProLeuThrValLeuLys 322
Db 7499 AACCTTACCTCCCTTACTGCGT 7479

RESULT 22
AB067197/c
ID AB067197 standard; DNA; 1163020 BP.
XX
XX AB067197;
AC
XX 29-AUG-2002 (first entry)
DT
XX
XX Listeria innocua contig DNA sequence #10.
XX

SQ Sequence 1233 BP; 468 A; 225 C; 241 G; 299 T; 0 other;

Alignment Scores:

Pred. No.:	0.41	Length:	1233
Score:	134.00	Matches:	62
Percent Similarity:	39.23%	Conservative:	37
Best Local Similarity:	24.60%	Mismatches:	111
Query Match:	8.02%	Indels:	42
DB:	24	Gaps:	10

US-10-018-706-2 (1-322) X AAL46626 (1-1233)

OY		7	GlnValIysGlnIleAspThrValSerLysIleAlaGlnArgIleAsnTrpArg	93		
Dd		562	CAAAAAAACAACAACAGCATTCGAAAAGAACAAGCAGAAAGCATATTTGC-----	6155		
OY		94	GluILeGIYHISIIleAmnLeuAnsSerSerrYrThnIleYrThngIyGlnTrpLeu	113		
Dd		616	-----CTGAATGACTCATATAAAATTTAGCCCTTGNATCAAGATTA-----	657		
OY		114	ThreutrpSerGIyspleuIlysvAlarGIluaGserIleSerSerGIyvalAsmThr	133		
Dd		658	-----TTGAATGCACCTAAAGCAAACGAACGACACTTGCGTCAAAATATCAACA	708		
OY		134	AlaHisThrProSerProValAlavalGInSerSeraTrgProProval--GlngInHis	152		
Dd		709	GCGAACAACAGCACACGCCAACGAACAAACGTGAAGAAGAGGACCTTGCTCAACGGCAA	768		
OY		153	ProAlavalGInIlysbProThrPrroProvalValvalIylsPvrothrProThrPro	172		
Dd		769	AAGCGTGAACAAAAACGAACA-----	789		
OY		173	ProvalValalGInIlnProAlaprovalAlapropovalThnrgualapropheAlatr	192		
Dd		790	-----TCAAAACCCTTATCAACCAACGTGCAAGAACGCCATTTCTTAAT	834		
OY		193	GlySerSerGIyval-----MetGlnPheaArgYrProvalGlYAlatPhasn	208		
Dd		835	AGTACAAGCGGTATTAGGGCGGCAAAAAACAATATCTCTTACAGATTTCGGTCA---	891		
OY		209	ProvalValalargArphheglYrThralatPhrValAlaGlySerThnValThnSerAnGIy	228		
Dd		892	----ATTITGGATCACTTTGGT-----TCTATCCAAGACGGCAAGTACGTGGAAAGGT	942		
OY		229	MetTrpPheSerGIyArgraspgIyAspleuIleasnAlaseranlAgIyThrValIle	248		
Dd		943	ATGTATTATGTGCATCAGACGGCAGCGCTGTAAAGCAATATCGACGTGGACGGCTTATT	100		
OY		249	GlnAlaAspHis---AsnMeCaspgIyAlaserIlevalIleglnHisThrasngIyPhe	267		
Dd		1003	TTAGCGGGAATTTAAATAGTATAGTTATATGTATATGTATTAACAACGGGAAACTGAT	106		
OY		268	ValSerSerYrIleHisIleYssAspaIagIvalIlysrhgIyAspThrValalgrThr	287		
Dd		1063	TTTAGATTATATATGGCTTCATCAACGGGTATCAGGAAGTTGGTCAGCTTGTTTCGCA	112		
OY		288	GIyGlnArGIleAlaserMetLyssaSn-----GlnProSerGIyAlalalePheGlu	305		
Dd		1123	GGGCGAGATTATGCTCAAGTACGAATATCAGGGGAATATATCAGCTTGCGCTTTAT---	117		
OY		306	PheArGIleserArghsnGIyValYrValaIsppro	317		
Dd		1180	TTTGATTTAGCCGTAAAGAACGCCAGTAAATATCA	1215		
RESULT 25						
ID	AAAL46624	standard; DNA;	1233 BP.			
AC	AAAL46624;					
DT	05-AUG-2002	(first entry)				
XX	H influenzae BMSB201 coding sequence #1.					
DE						

XX	BASB201; otitis media; pneumonia; sinusitis; nosocomial infection,
KM	auditive nerve damage; delayed speech learning; vaccine,
KM	antibacterial; auditory; antiinflammatory; gene; ds.
XX	
OS	Haemophilus influenzae.
XX	

PH	Key	Location/Qualifiers
FT	CDS	1.1233
FT		/*tag= a
FT		/product= "BASB201"
..		

PN WO200230967-A2

PD 18-APR-2002
yy

05-OCT-2001; 2001WO-EP11561.

PR 13-OCT-2000; 2000GB-0025169.

PA (SMK) SMITHKLINE BEECHAM BIOLOGICALS.
XX

Pl. Thonnard J,
XX.

DR WFL; 2002-426261/45
DR P-PSDB; AA017656.

AA PT	New isolated non-
----------	-------------------

PT as otitis media, delayed speech learning and inflammation of middle ear.

PS Claim 13; Page 87; 90pp; English

The present invention provides the protein and coding sequences of several versions of the Babs201 protein from non-typable *Haemophilus influenzae*. These can be used in the production of vaccines against *H. influenzae* infection, which can cause otitis media in infants and children, pneumonia in elderly, sinusitis, nosocomial infections, or invasive diseases, chronic otitis media with hearing loss, fluid accumulation in the middle ear, auditory nerve damage, delayed speech learning, infections of the upper respiratory tract and inflammation of the middle ear. The present sequence is a version of the Babs201 coding sequence of the invention.

Sequence 1233 BP; 466 A; 224 C; 239 G; 304 T; 0 other;

Alignment Scores:

Pred. No.:	0.688	Length:	1233
Score:	131.00	Matches:	62
Percent Similarity:	38.89%	Conservative:	36
Best Local Similarity:	2.60%	Mismatches:	112
Query Match:	7.84%	Indels:	42
DB:	24	Gaps:	10

US-10-018-706-2 (1-322) X AAL46624 (1-1233)

OY	7	GlnValIlyGInGIyAspPthrValSerIysrLleAlaGlnArgIyGrGLyLeuAsnTrpArg	93
Db	562	CAAAAATAACAAACAAGCATTTGCCAAMAAACACAGCAAGAATCAA-----	6099
OY	94	GluILeGIyHISlIAsnAsnLeuAsnSerSerIyrThrIleIyrrThnGIyGIntPrIeu	113
Db	610	-----TTCAGCGTAATGCATCTCAATAAAATTTAGCCGTATGCAGATAA-----	657
OY	114	ThrLeutprSerGIyAspLeuLyValrGlulArSerIlIeserSerGIyAlaAsnThr	133
Db	658	-----TTGAATGCATTAAGAACAAACAGCAGACTTGTCGACGAATTTCAACGA	708
OY	134	AlaHisThrProSerTrpValAlaValGlnSerSerArpProproVal--GInGlnHis	152
CY	709	GCTGAACAAGCACGGCAACCAAGAAAAAGGTGAAGAAGAGCACTTGCTTCAACGCCAA	768

Oy		153	ProValAlaIGlnLysPrcThProProValValValLysLySPrOThPrOThPrO	172
Db		769	AAAGCTGAAGAAAACGACA-----	789
Oy		173	ProValValGInGIInPrOAlaPrOValAlaPrOProValThrGlUAlaPrOheAlaThr	192
Db		790	-----TCAAGCCTTATCACCAACTGTGGCAAGAAGCCAACTTACTTAAT	834
Oy		193	GlySerSerGIYValI-----MetGInPheArgTYrProValIGLYAlaThraSn	208
Db		835	AGTACAAACCGGTTTGAGGGCGGCMAAAAAACAATATTCCTTACACAGTTTCTGGTTCA---	891
Oy		209	ProValValAArgArXpHeGLYThrLaThrValAlaIGlySerThValIThrSerangly	228
Db		892	--ATTTCGATCATCTTTGGT-----TCATTTCAAGACGGCAAGTACGTTGGAAAGST	942
Oy		229	MetTrPheSerGIYArgAspGIYAspLeuIleAlaSerAsnAlaGIYThrValIle	248
Db		943	ATGCTAATATGGCGCATACGACGACGACGCCCTGTAAAGCAATTGCTGCTGGACGCTCATTT	1002
Oy		249	GIAlaAspHis--AsnMetAspGIYAlaSerIleValIleGInHisIThraSnglyPhe	267
Db		1003	TTAGGGGAGATATTATAAGTGTATGTATAGTTATTGTTAAACACGGCGAAACTGAT	1062
Oy		268	ValSerSerTYrIleHisIleLysAspAlaGInValYsthrGIYAspThValArgthr	287
Db		1063	TTAAGTTATATATGGCTTCATCAACCTGTATCAGGAAGTAGTGTCACCTTGTTTCAGCA	1122
Oy		288	GLYGIAlaArgIleAlaSerMetLysasn-----GInProSerGIYAlaAlaLeuPheGIU	305
Db		1123	GGGCAAGCTTATTCCTCAAGTGAATAATCAGGGGAAAATATACGTTCTGCGCTTAT---	1179
Oy		306	PheArgIleSerArgAsnglyValTYrValaspRo	317
Db		1180	TTTGSTATTTAGCCGTAAAGAAACGCCAGTAAATCCT	1215
RESULT 26				
AAQ51086				
ID	AAQ51086	standard; DNA;	3646 BP.	
XX	AAQ51086;			
AC				
XX				
DT	19-APR-1994	(first entry)		
XX				
DE	Plasmid pCRR28	which encodes the lppB protein.		
XX				
KW	Haemophilus somnus;	immunogenic; haemolytic; lppB; lppC;		
KM	thromboembolic meningococcalitis; septicemia; arthritis;			
KM	pneumonia; lppA gene; Pasteurella haemolytica; ds.			
OS	Haemophilus somnus.			
OS	Pasteurella haemolytica.			
XX				
FH	Key	Location/Qualifiers		
FT	CDS	1..2772		
FT		/*tag= a		
FT		/note= "leukotoxin gene"		
FT	CDS	2773..3573		
FT		/*tag= b		
FT		/note= "lppB peptide (split)"		
XX				
PN	WO93121323-A.			
PD	28-OCT-1993.			
XX				
PF	05-APR-1993;	93WO-CA00135.		
XX				
PR	09-APR-1992;	92US-0865050.		
PR	04-JUN-1992;	92US-0893424.		
PR	29-MAR-1993;	92US-0893426.		
PR	29-MAR-1993;	93US-0038287.		
PR	29-MAR-1993;	93US-0038288.		
PR	29-MAR-1993;	93US-0038719.		

```

XX      (UTSA-) UNIV SASKATCHEWAN.
PA      Harland RJ, Pfeiffer CG, Pontarollo RA, Potter AA;
XX      Rioux C, Theisen M;
PI      WPI, 1993-351733/44.
XX      P-P8DB; AAR42385-6.
DR      Haemophilus somnus immunogenic proteins used in vaccines -
XX      PT selected from haem-in-binding protein, haemolysin, LppB and LppC,
XX      PT and corresp. DNA
XX      Claim 8; Fig 11, 119p; English.
PS      The lppB gene encoding lppB was expressed in E. coli as a fusion to
CC      the Pasteurella haemolytica leukotoxin gene lxtA coded for by plasmid
CC      pAA352. The lppB gene fragment was taken from pMS11. The lppB can be
CC      used in vaccines for preventing or treating H. somnus infections, which
CC      cause thromboembolic meningitis-encephalitis, septicemia, arthritis and
CC      pneumonia in vertebrates.
XX      See also AAO51080-5.
XX      SQ Sequence 3646 BP; 1257 A; 646 C; 749 G; 994 T; 0 other;
Alignment Scores:
Pred. No.:      2.55      Length:      3646
Score:          131.00     Matches:      50
Percent Similarity: 44.71% Conservative: 43
Best Local Similarity: 24.04% Mismatches: 77
Query Match:     7.84%     Indels:      38
DB:              14       Gaps:        7

```

```

US-10-018-706-2 (1-322) x AAQ51086 (1-3646)
QY 73 TyrGlnValVlysgInGlyAspThrValSerIstIleaglArgIyrgIyLeuAsnTrp 92
Db 3091 TACAAAGTACGCCAAAGCGATACCATGTTCTTATGCTTATATTCAGGATGGATATTA 3150
QY 93 ArgGluIleGlyHisIleAsnAspLeuAsnSerSerIyThrIleIyThrGlyGlnTrp 112
Db 3151 AAAGATTTGGCCACACTTAAATATATGTCTGAGCCATATCTCTGAGATTGGACAACTA 3210
QY 113 LeuThrIeuTrpSerGly-----AspLeuIySValArgGluArgSerIleSerSer 129
Db 3211 TTGAATATTCGAATATATATTCGCCATAGCAATATGATATACCAACAGACATTAATGAA 3270
QY 130 GlyValLeuThrIleHisIleThrProSerProValAlaValGlnSerSerArgProProVal 149
Db 3271 TCAGAGGTGACACAAATATACAGTCAATGAG---ACATGGAATGCTAATTAACCAACAAT 3327
QY 150 GlnGlnHisProAlaValGlnIyProThrProProValValValIyIyIyProThr 169
Db 3328 GAACAATATGAACCCGTTGCTACACCAACATTTCAACAATGCCAATCAATATTA--- 3381
QY 170 ProThrProProValValGlnGlnProAlaProValAlaProProValThrGluAlaPro 189
Db 3382 ---ACACCTCCAGCGACCTCAATATATGCTTGATTTGGCCA----- 3420
QY 190 PheAlaThrGlySerSerGlyValMetGlnPheArgTyProValGlyAlaThrAsnPro 209
Db 3421 -----ACAAATGCA 3429
QY 210 ValValArgArgPheGlyThrAlaThrValAlaGlySerThrValThrSerAsnGlyMet 229
Db 3430 AAAATATATCA---GGAATTTTCCAGTGGCTGATGAGGC-----AATTAAGGATTT 3477
QY 230 TrpPheSerGlyArgAspGlyAspLeuIleAsnAlaSerAsnAlaGlyThr--ValIleG 249
Db 3478 GATATTGACGGTTCGCGGACAGACGTGTATATGCACAGCTGATGAGACGCAATTTGAT 3537
QY 249 IuAlaAspHisAsnMetAspGlyAlaSer-----IleValIleGlnHisThrAnGlyP 267

```

Db 3538 ATCCGAGACGCTTACGTGATGTAATTAATTATTATTAAACATGACAGTT 3597
QY 267 heValSerSerTyrIleHis 273
DB 3598 ATTTAAGTGTCTATGACAT 3617
RESULT 27
AA06135
ID AA06135 standard; DNA; 1929 BP.
XX AA06135;
XX
XX 01-APR-1996 (first entry)
XX
XX Lysin coding sequence.
XX
XX Lysin; autolysis; culture; lactic acid bacteria; fermentation;
XX cheese, foodstuffs; induction; ds.
XX
XX Lactococcus lactis (Strain MG1363).
OS
FH Key Location/Qualifiers
FT CDS 178..1491
FT /tag= a
FT /product= Lysin.
FT complement (1542..1824)
FT /*tag= b
XX
XX MO9531561-A1.
XX
XX 23-NOV-1995.
XX
XX 12-MAY-1995; 95WO-NL00170.
XX
XX 12-MAY-1994; 94EP-0201353.
XX
XX (UNIL) QUEST INT BV.
XX
XX Buier G, Kok J, Ledebroer AM, Venema G,
XX
XX WPI; 1996-010946/01.
XX P-P8DB; AAR85285, AAR85288, AAR85289.
XX
XX
XX Lysis of a culture of lactic acid bacteria in, e.g. cheese
XX production - by in situ prodn. of an auto-lysin, regulated by an
XX inducible promoter.
XX
XX Claim 14; Page 51-54; 103pp; English.
XX
XX In situ production of a homologous autolysin or a heterologous
XX autolysin from a food grade Gram positive bacteria, can be used in
XX a process for the lysis of a culture of lactic acid bacteria. The
XX process can be used in the manufacture of products containing
XX cultures of lactic acid bacteria e.g. cheese, where the culture is
XX lysed following the completion of fermentation. The enhanced
XX induction of the autolysin is performed some hours after the
XX fermentation is finished. No extra lysin needs to be added and the
XX lysin does not need to be isolated or encapsulated. The time of
XX lysis can be precisely controlled.
XX
XX
SQ Sequence 1929 BP; 619 A; 378 C; 321 G; 611 T; 0 other;
Alignment Scores:
Pred. No.: 1.29 Length: 1929
Score: 130.50 Matches: 78
Percent Similarity: 37.93% Conservative: 43
Best Local Similarity: 24.45% Mismatches: 107
Query Match: 7.81% Indels: 92
DB: 17 Gaps: 15
US-10-018-706-2 (1-322) x AA06135 (1-1929)
QY 26 ThyrThrCysIleLeuAlaGlyCysAlaSerLysProThrTyrAsnSerThr-----42

Db 727 ACTGCTGCTTTACGGGACAGATATGCTACGACCAGCAATGAGGCTTACGTATGCC 786
QY 43
Db 787 ATTATTTCTCAATATTAATTGACTGCTTTGACGAGGCTTCTTCACTGGAATATCTAT 846
QY 53 SerGlyGlyLeuAlaIleGlySerGlnValIleThrAspSerGlnGlyValProAsn-- 71
Db 847 TCTGGTGGC-----TCGACAAACCAAAATTCAGAAATTAATTTCTGGAACCAATGCG 897
QY 72 -----ArgTyrGlnValIleGlnGlnIleAspThrValSerLysIleAlaGlnArgTyr 88
Db 898 AGTTCAACTACTTAATACCGTCAMATCTGTGATACTCTTGGGGAATCTCACAAAGATAT 957
QY 89 GlyLeuAsnTrpArgLuuIleGlyHisIleAsnLeuAsnSerSerTyrThrIleTyr 108
Db 958 GGAATTTAGTGTGCGCTCAAAATTCAAAGTCGAAATATCTTAAGATACC---ATTATCTAC 1014
QY 109 ThrGlyGlnTrpLeuThrLeuTrpSerGlyAspLeuLysValArgGluArgSerIleSer 128
Db 1015 ATTGCTCAAAAACCTGTACTGACAGATTCA-----GCTTCTTCACAAAATTC 1062
QY 129 SerGlyValAsnThrAlaHisThrProSerProValAlaValGlnSerSerArgProPro 148
Db 1063 GGTGGTTCAAAACAATTC-----GCAGACGACTACCAAC 1098
QY 149 ValGlnGlnHisProAlaValGlnLysProThrProProValValValValLysLysPro 168
Db 1099 ACTTCTGACACCTGCA-----AACCACCTTCACAAACACTGTTAAGTTAATCC 1152
QY 169 ThrProThr-----ProProValValGlnGlnProAlaProValAlaProProValThr 186
Db 1153 GGAATATCCCTTGGGCGCTTCAAGTAAATATTAATAACTAGTATGCTCAATGAAATG 1212
QY 187 GluAlaProPheAlaThrGlySerSerGlyValMetGln-----PheArgTyrPro 203
Db 1213 TGGAAATCATTTAAGTTCAGATACATTTATGTCAMAAATCTTATGTTTCACAATCT 1272
QY 204 ValGluAlaThrAsnProValValArgArgPheGlyThrAlaThrValAlaGlySerThr 223
Db 1273 GCTGCTGCTTCAAAATTCCT-----TCGACAGGTTTCAGGCTCAACT 1311
QY 224 ValThrSerAsnGlyMetTrpPheSerGlyArgAspGlyAspLeuIleAsnAlaSerAsn 243
Db 1312 GCTACCAATTAAC-----TCAAAC 1329
QY 244 AlaGlyThrValIleGlnAlaAspHisAsnMetAspGlyAlaSerIleValIleGlnHis 263
Db 1330 TCG----- 1332
QY 264 ThrAsnGlyPheValSerSerTyrIleHisIleLysAspAlaGlnValLysThrGlyAsp 283
Db 1333 ACTTCTTCTAATCAATTAATGCTCAATTCAT-----AGGTCGTTAAAGGAT 1380
QY 284 ThrVal-----ArgThrGlyGlnArgGlyIleAspMetLys-----Asn 296
Db 1381 ACTTCTGGGACCTTGGCAAAATTCGCAAGCCCAATTCCTTCAATCAAGGCTTGAAAT 1440
QY 297 GlnProSerGlyAlaAlaLeuPheGlnPheArgIleSerArgAsnGlyValIleVal 315
Db 1441 CATTTATCTTACGCACTACTAT-TTATTTGGTCAGTATCTCAAGATTAATAATTAATT 1496
RESULT 28
AA023454
ID AA023454 standard; cDNA; 4093 BP.
XX
XX AAA23454;
XX
XX
XX 19-JUN-2000 (first entry)
XX
XX cDNA encoding human secreted protein vb22_1, SEQ ID NO:63.
XX
XX
XX

Human; secreted protein; cancer; tumour; cardiovascular disorder;
blood disorder; haemophilia; autoimmune disease; diabetes; inflammation;
infection; fungal; bacterial; viral; HIV; allergy; arthritis;
neurodegenerative disease; asthma; contraceptive; ss.

Homo sapiens.

Key Location/Qualifiers

CDs 1048..3729

FT /*tag= a

FT /product= "Human secreted protein vb22_1"

FT 152..1006

FT /*tag= b

FT /product= "Clone vb22_1 ORF2"

PN WO200011015-A1.

XX 02-MAR-2000.

XX 24-AUG-1999; 99WO-US19351.

XX 24-AUG-1998; 98US-0097638.

XX 24-AUG-1998; 98US-0087659.

XX 09-SEP-1998; 98US-0099618.

XX 28-SEP-1998; 98US-0102092.

XX 25-NOV-1998; 98US-0109978.

XX 23-DEC-1998; 98US-0113645.

XX 23-DEC-1998; 98US-0113646.

XX 23-AUG-1999; 99US-0379246.

XX (ALPH-) ALPHAGENE INC.

XX Valenzuela D, Yuan O, Hoffman H, Hall J, Rapiejko P;

XX WPI, 2000-224657/19.

XX P-PSDB; AAY95012, AAY95030.

XX New secreted or transmembrane proteins and polynucleotides encoding

XX them, useful for treating neurodegenerative disorders, autoimmune

XX diseases and cancer -

XX Claim 72; Page 321-322; 357pp; English.

XX The invention relates to 40 human secreted proteins (AAY94981-Y95020),

XX and cDNA sequences encoding them (AAA23423-A23462). The secreted

XX proteins of the invention include those that are thought to be only

XX partially secreted, i.e., transmembrane proteins. The proteins of the

XX invention may exhibit one or more activities selected from the following:

XX cytokine activity; cell proliferation; differentiation; immune

XX modulation; haematopoiesis regulation; tissue growth activity;

XX activin/inhibin activity; chemotactic/chemokinetic activity; haemostatic

XX and thrombolytic activity; anti-inflammatory activity; and tumour

XX inhibition activity. The proteins may be administered to patients as

XX vaccines, and the nucleotides may be used as part of a gene therapy

XX regime. Diseases or conditions that may be treated using the proteins or

XX nucleotides of the invention include autoimmune diseases; genetic

XX disorders; haemophilia; cardiovascular diseases; cancer; bacterial,

XX fungal and viral infections, especially HIV; multiple sclerosis;

XX rheumatoid arthritis; pulmonary inflammation; Guillain-Barre syndrome;

XX insulin dependent diabetes mellitus; and allergic reactions such as

XX asthma and anaemia. They may also be used for treating wounds, burns,

XX ulcers, osteoporosis, osteoarthritis, periodontal diseases, Alzheimer's

XX disease, Parkinson's disease, Huntington's disease and amyotrophic

XX lateral sclerosis (ALS). Proteins with activin/inhibin activity may

XX additionally be useful as contraceptives. Nucleic acid sequences of the

XX invention may be used in chromosome mapping, and as a source of

XX diagnostic primers and probes. The present sequence represents cDNA

XX encoding one of the 40 proteins of the invention.

XX Sequence 4093 BP; 1213 A; 926 C; 928 G; 1026 T; 0 other;

XX Alignment Scores:

XX Pred. No.: 3.19

XX Length: 4093

Score: 130.50 Matches: 59
Percent Similarity: 40.09% Conservative: 26
Best Local Similarity: 27.83% Mismatches: 85
Query Match: 7.81% Indels: 42
DB: 21 Gaps: 10

US-10-018-706-2 (1-322) x AAA23454 (1-4093)

QY 132 AsnThrAlaHisThrProSerProVal-----AlaValGlnSerSerArgProPro 148
DB 494 TCGACCGTGGCCCGCGCCATCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 553
QY 149 ValGlnGlnHisProAlaValGlnLysProThrProProValAlaValLys----- 166
DB 554 GAGACGACGACGCTCCGCGCGCGCTCCCTCTCCCTCCCTCCCTCCCTCCCTCCCTCC 613
QY 167 --LysProThrProThrProProProValValGlnGlnProAlaProAlaProProVal 185
DB 614 GCAGAGCGCGTGTGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 664
QY 186 Thr---GluAlaProPheAlaThrGlySerSerGlyValMetGln-----PheArg 201
DB 665 ACCCG 724
QY 202 TyrProValGlyAlaThrAsnProValAlaArgArgPheGlyThrAlaThrValAlaGly 221
DB 725 CTTCCT---GCTGATCTGAGCGCTGTGATAGCTCC----- 757
QY 222 SerThrValThrSerAsnGlyMetTrpPheSerGlyArgAspGlyAspLeuIleAsnAla 241
DB 758 -----TCTGCAGAAAATATGACTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 808
QY 242 Ser-----AsnAlaGlyThrValIleGlnAlaAspHis 252
DB 809 GGTCAAGAGATTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 868
QY 253 AsnMetAspGlyAlaSerIleValIleGlnHisThrAsnGlyPheAlaSerSer-TyrIle 272
DB 869 CTTCTCAGCGCGCTTCTTCAAGAACATGAATACCTGTGATATTGTCTCAACGATTA 928
QY 272 e-----HisIleLysAspAlaGlnValLysThrGlyAspThrValArgThrGly 288
DB 929 CCCACTGAAGAACATCTTCAAGAAAATGTCTAGTGAAG-----CTTAAAGAGGTC 979
QY 288 yGlnArgIleAlaSerMetLysAsnGlnProSerGlyAlaAlaLeuPheGluPheArgIle 308
DB 980 TCAGAGAGAGGCAAACTCTACTATGATAGATTAACAGAGTTTCAGAAATTAGAAAT 1039
QY 308 eSerArgAsnGlyValTyrValAspProLeuThr 319
DB 1040 ACTCAGAAATGGATCATCGTTCACTGCTCTCTCC 1073

RESULT 29

ABL50562/C

ID ABL50562 standard; DNA; 11115 BP.

XX ABL50562;

XX 18-JUN-2002 (first entry)

XX Micromonospora carbonacea evernimicin locus nucleotide contig 8.

XX Micromonospora carbonacea; antibiotic; evernimicin; biosynthesis;

XX gene cluster; genetic manipulation; contig; gene; ds.

XX Micromonospora carbonacea.

XX WO200155180-A2.

XX 02-AUG-2001.

XX 29-JAN-2001; 2001WO-CM00128.

CC from the vep gene cluster and module 7 from the Streptomyces ty1P.
CC gene cluster has polyhydroxyalkanoate (PHA) monomer synthase
CC activity and can be used for PHA prodn. in host (esp. insect) cells
CC for use as a biodegradable polymer.

XX Sequence 15872 BP; 2093 A; 6313 C; 5503 G; 1963 T; 0 other;

Alignment Scores:

Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
Score:	16.4	15872	81	43	115	18
Percent Similarity:	130.50					
Best Local Similarity:	38.75%					
Query Match:	25.31%					
	7.81%					

US-10-018-706-2 (1-332) x AAT68715 (1-15872)

```

Qy 34 AlaSerlyProthryrYarnSerThSerGlySerHisatgThSerGlySer 53
Db 3195 GCGTCGCGCCCTC-----TCGCGACGAGCCCGCACAGGTCTCTCGGGGAG 3151
Qy 54 -----GlyGlyLeuAlaIleGlySerGlnValIle 63
Db 3150 GTGACCGCCCGGGGTAGCGGACGATGCCATGATGCGATGCGCTCGCGGTGTG 3091
Qy 64 ThrAsp-----SerGlnGlyValProAsnArgTyrGlnValIleGlnGlyAspThrVal 81
Db 3090 TCGCGCGGGGTGCGGGGCGGTATCCGTCGCGCCATCCGTCGCGCCGCGCGGTG 3031
Qy 82 SerlyS-----IleAlaGlnArgTyrGlyLeuAsnTyrPargGlu----- 94
Db 3030 AGCAGGTCGCGCCAGTGGGCGCGGAGCGCGCGCGTGGGTGTGAGAGACAGTCCG 2971
Qy 95 ILGlyHisIleAsnLeuAsnSerSerTyrThrIleTyrThrGlyGlnTrpLeuThr 114
Db 2970 CTGGGCGAGCGCACTCCGTCGTCTGTCAGACGCGCTTCGCGAC----- 2926
Qy 115 LeuTrpSerGlyAsp-----LeuLysValArgGluArgSer 126
Db 2925 -----TCGACGACATAGAGAGAGTCCAGCCGACCTCTTGAAGTGTAGCCGATTCG 2872
Qy 127 IleSerSerGly-----ValAsnThrAlaHisThrProSerPro 139
Db 2871 ACGGGCGTGGGTGTCTACTCGAGAGCGGCGGACGATGTCGCGACGCGCTCGGTC 2812
Qy 140 ValAlaValGlnSerSerArgProProValGlnGlnHisProAlaValGlnLysProThr 159
Db 2811 GTCTCGTGGGAGCGCGCGCATACCG-----CCACCCGACCGCGCGCTCTCG 2764
Qy 160 ProProValAlaValLysLysProThrProThrProProValValGlnGlnProAla 179
Db 2763 CCC-----TCGCTCTCGCTCGCTCGCGCCACGTCACACCGCGCGCGG 2722
Qy 180 ProValAlaProProValThGlnAlaProPheAlaThrGlySer----- 194
Db 2721 CCGGTGCGCCGACGCGGTGCGCGCATGCGCGCGCGCGCGCGCGCGCGCGCGG 2662
Qy 195 -----SerGlyValMetGlnPheArgTyrPro-----ValGlyAlaThrAsnProValVal 211
Db 2661 CCGTCGACACAGTGGCTGCTGCGCTGGAAGCGTAGTGTGCGACGCGCACCTGACCGTGTG 2602
Qy 212 ArgArgPheGlyThrAlaThrValAlaGlySer-----ThrValThr 225
Db 2601 CCGGTGCTCCCGTCCGCGCGCGGTCTCAGTCAGTCGTCGCGCGCGCGCGCGCGCGGTG 2542
Qy 226 Ser-----AsnGlyMetTyrPheSerGlyArgAspGlyAspLeuIleAsnAlaSerAsn 243
Db 2541 AGTGGCGGAGACGACGATGGGGGCTCGGGGCGCGCC-----TTGCGCAGG 2497
Qy 244 AlaGlyThrValIleGlnAlaAsp-----HisAsnMetAspGlyAlaSerIleValIleGln 262
Db 2496 GCGGAGACCGCCCGTGGCGCTCTGTGTGCGGTACGAGAGTCCCGCGCATCGCGAGACAG 2437

```

```

Qy 263 HisThrAsnGlyPheValSerSerTyrIleHisIleLysAsp---AlaGlnValIleThr 281
Db 2436 ACCCGTCGAGACCGAGCTCAGAGAGTGTGCGGCGCCGATTCCTCCAGGTGTGTCG 2377
Qy 282 GlyAspThrValArgThrGlyGlnArgGlyLeaSerMetLysAsnGlnProSerGlyAla 301
Db 2376 GCGTCGAGAGACGATGCGGCGCTGCGG-----ACGTGTCACACGACTACTCGGATCG 2323

RESULT 31
ID AA287283/c
ID AA287283 standard; DNA, 15872 BP.
XX AA287283;
XX
XX 05-JUN-2000 (first entry)
XX
XX S. venezuelae vep ORF 1, SEQ ID NO:1.
XX
XX Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
XX neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
XX biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
XX chronic obstructive pulmonary disease; respiratory inflammation;
XX hypercholesterolemia; crop protection agent; ds.
XX
XX Streptomyces venezuelae ATCC15439.
XX
XX
XX Key Location/Qualifiers
XX CDS 20..13912
XX
XX /tag= a
XX /product= "vep ORF 1 amino acid sequence #1 (AAV77177)"
XX CDS 14056..14151
XX /tag= b
XX /product= "vep ORF 1 amino acid sequence #3 (AAV77199)"
XX CDS 14167..15827
XX /tag= c
XX /product= "vep ORF 1 amino acid sequence #2 (AAV77178)"
XX
XX WO200000620-A2.
XX
XX 06-JAN-2000.
XX
XX 25-JUN-1999; 99WO-US14398.
XX
XX 26-JUN-1998; 98US-0105537.
XX
XX (MINU ) UNIV MINNESOTA.
XX
XX Sherman DH, Liu H, Xue Y, Zhao L;
XX
XX WPI; 2000-160679/14.
XX P-PSDB; AAY77177, AAY77178, AAY77199.
XX
XX Desosamine and macrolide biosynthetic gene clusters, useful for, e.g.
XX synthesis of methymycin and pikromycin -
XX
XX Example 3; Figure 23; 438pp; English.
XX
XX The invention relates to an isolated and purified nucleic acid segment
XX comprising a desosamine biosynthetic gene cluster, a fragment of its
XX biologically active variant, where the nucleic acid sequence is not
XX derived from the eryC gene cluster of Saccharopolyspora erythraea or
XX Streptomyces antibioticus. The invention also relates to a macrolide
XX biosynthetic gene cluster, or fragments thereof. The macrolide
XX biosynthetic gene cluster encodes proteins which synthesise methymycin,
XX pikromycin, neomethymycin, narbomycin or a combination of these
XX compounds. Recombinant or augmented cells comprising the desosamine
XX and/or macrolide biosynthetic gene clusters are useful for the
XX production of biologically active macrolides. The macrolide biosynthetic
XX proteins are useful for synthesis of methymycin, pikromycin,
XX neomethymycin and narbomycin. The alternative termination of polyketide
XX synthesis may be useful to prepare novel antibiotics and
XX polyhydroxyalkanoate (PHA) monomers. The compounds produced by the
XX recombinant host cells are useful as biopolymers, e.g., in packaging or

```

CC biomedical applications, to engineer PNA monomer synthases or to prepare
 CC biologically active agents, such as chemotherapeutics,
 CC immunosuppressants, agents to treat asthma, chronic obstructive pulmonary
 CC disease as well as other diseases involving respiratory inflammation,
 CC cholesterol-lowering agents or macrolide-based antibiotics which are
 CC active against a variety of organisms, e.g. bacteria, including
 CC multi-drug resistant pneumococci and other respiratory pathogens, as well
 CC as viral parasitic pathogens, or as crop protection agents (e.g.,
 CC fungicides or insecticides) via expression of polypeptide genes.
 CC The present sequence represents a Streptomyces venezuelae ATCC 15413 DNA
 CC sequence, designated vep ORF 1 in the specification, which actually
 CC contains 3 open reading frames, which encode proteins AY7117-17118 and
 CC AY7119. The vep ORF 1 protein is defined in the specification as a PNA
 CC monomer synthase.

SO Sequence 15872 BP, 2088 A, 6304 C, 5513 G, 1967 T, 0 other;

Alignment Scores:

Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:
Score:	16.4	81	43	115	81
Percent Similarity:	130.50				
Best Local Similarity:	38.75%				
Query Match:	25.31%				
DB:	21	Gaps:	18		

US-10-018-706-2 (1-322) x AA287283 (1-15872)

```

QY 34 AlaSerIysProThrTyraSenSerThrSerGlySerHisArgThrSerGlySer 53
DB 3195 GCGTCGCGCGCC-----TCGCGCAGCAGCGCGCCACAGCTCTCGGGGAG 3151
QY 54 -----GlyGlyLeuAlaIleGlySerGlnValIle 63
DB 3150 GTGACCGCGCGGGGTAGCGCGCAGCCATGCGCATGCGCATGGCTCTCGGGGTG 3091
QY 64 ThrAsp-----SerGlnGlyValProAlaArgTyrGlnValIleGlnValAspThrVal 81
DB 3090 TCCGCGGGGGTGCGGCGGTATCCCTGCGCGCATCGGCTCTCGCGCGCGGGT 3031
QY 82 SerIys-----IleAlaGlnArgTyrGlyLeuAsnTyrArgIle 94
DB 3030 AGCAGCTCCCGCAGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2971
QY 95 IleGlyHisIleLeuAsnLeuAsnSerTyrThrIleTyrThrGlyGlnTyrLeuThr 114
DB 2970 CTGGCGAGCGCGAGCTCCGCTGCTGTCGACGAGCGCGTCCGAGC----- 2926
QY 115 LeuTyrSerGlyAsp-----LeuIysValArgGlnArgSer 126
DB 2925 -----TCGACGACATGAGGAGATGCAAGCCAGCTCCTTGAAGGTAGCGCGAGTTG 2872
QY 127 IleSerSerGly-----ValAsnThrAlaHisThrProSerPro 139
DB 2871 ACGCGGTCGGGTCTGCTACTGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 2812
QY 140 ValAlaValAlaGlnSerSerArgProProValGlnGlnHisProAlaValGlnIysProThr 159
DB 2811 GCTCTGTCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2764
QY 160 ProProValValValValIysLysProThrProThrProProValValGlnIleProAla 179
DB 2763 CCC-----TCGCGCTGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCG 2722
QY 180 ProValAlaProProValThrGlnAlaProPheAlaThrGlySer----- 194
DB 2721 CCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2662
QY 195 ---SerGlyValMetGlnPheArgTyrPro-----ValGlyAlaThrAsnProValVal 211
DB 2661 CCGTGAACCGAGTGCCTGCGCTGCGAGAGCGTACGTGGAGGAGGACCGCTGACCG 2602
QY 212 ArgArgPheGlyThrAlaThrValAlaGlySer-----ThrValThr 225
DB 225 -----ThrValThr 225

```

```

DB 2601 CCGGTCTCCCGTGGCGCGCGGTCCATGTCACGTCTGCGCCCGGACGACGAGTGTG 2542
QY 226 Ser-----AsnGlyMetTyrPheSerGlyArgAspGlyAspLeuIleAsnIleSerAsn 243
DB 2541 AGTGGCGGACGACGAGCTGGGGCTCCGGCGCGCC-----TTGGCGCAGG 2497
QY 244 AlaGlyThrValIleGlnAlaAsp---HisAsnMetAspGlyAlaSerIleValIleGln 262
DB 2496 GCGGAGACCGCGCTGCGCGCGCTCTGTCGCGTACGAGTCCGCGCGCATCGCGAGCAG 2437
QY 263 HisThrAsnGlyPheValSerSerTyrIleHisIleLysAsp---AlaGlnValIysThr 281
DB 2436 ACCCGCTGGGACCGAGCTCCAGAGAGGTGTCGCGCGCGCGCGCGCGCGCGGTACG 2377
QY 282 GlyAspThrValArgThrGlyGlnArgIleAlaSerMetLysAsnGlnProSerGlyAla 301
DB 2376 GCGTCGAGGAGCGTACGCGCGCTGCGG-----ACCTGTCTCACCACGATCTCGGATCG 2323
RESULT 32
ABA90521
ID ABA90521 standard; DNA; 2365589 BP.
XX
XX ABA90521;
XX
XX 16-MAY-2002 (first entry)
XX
XX Genomic sequence of Lactococcus lactis IL1403.
XX
XX Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese; ds.
XX
XX Lactococcus lactis IL1403.
XX
XX PN FR2807446-A1.
XX
XX PD 12-OCT-2001.
XX
XX PF 11-APR-2000; 2000FR-0004630.
XX
XX PR 11-APR-2000; 2000FR-0004630.
XX
XX PA (INRG ) INRA INST NAT RECH AGRONOMICUE.
XX
XX PI Bolotine A, Sorokine A, Renault P, Ehrlich SD;
XX
XX DR WPI; 2002-043418/06.
XX
XX PT New nucleotide sequence useful in the identification of Lactococcus
XX lactis and related species -
XX
XX PS Claim 1; SEQ ID 1; 2504bp; French.
XX
XX
XX The present invention is related to a Lactococcus lactis nucleotide
XX sequence (ABA90521) and related proteins (ABR53300-ABR55621). The
XX nucleic acid sequence is useful in the detection and/or amplification of
XX related acid sequence, particularly to identify Lactococcus lactis or
XX CC biosynthesis or biodegradation of the invention are useful for the
XX CC invention helps research in lactic bacteria, particularly useful in the
XX CC production of yogurt and cheese.
XX Note: The sequence data for this patent is based on equivalent patent
XX WO20017734 (published 18-OCT-2001) which is available in electronic
XX format directly from WIPO at ftp.wipo.int/pub/published_pat_sequences.
XX
XX SQ Sequence 2365589 BP, 765914 A, 415261 C, 420487 G, 763927 T, 0 other;

```

Alignment Scores:

Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:
Score:	6.89e+03	79	2365589		
Percent Similarity:	130.50				
Best Local Similarity:	35.74%				
Query Match:	24.76%				
DB:	24	Gaps:	16		

```

US-10-018-706-2 (1-322) x ABA90521 (1-2365589)
Qy 3 ValThrIleAlaIleAsnSerGlnAsnGlnProIleValArgLeuGlyLeuIlePhe 22
Db 1065868 GTACCGG---ATTAACGCTACTGGCAATCTGGGCATAGGCATTTGGGCTGGT-TTC 1065923
Qy 23 GlyValIleThrThrCysIleLeuAlaGlyCysAla---SerLysProThrTyrAsnSer 41
Db 1065924 TCTTAACAAAGCGGCGCGCATTTACTCGGAATGACAGGGGTAATCTGATGCCAATCCA 1065983
Qy 42 ThrSerGlySerGlySerHisArgThrSerGlySerGlyGlyLeuAlaIleGlySerGln 61
Db 1065984 ACGGCTGATGAGGCGC-----GCGGACGCTCTCGATTCGGTTATGAGGCTTTGG 1066031
Qy 62 ValIleThrAspSerGlnGlyValProAsnArg-----TyrGlnValLys----- 76
Db 1066032 CAATGACGAGATAGTTACGGGCTAGCTCTGACGCTGTTTATATGATTAACCTCATGACA 1066091
Qy 77 -----GlnGlyAspThrValSerLysIleAlaGlnArgTyrGlyLeu 90
Db 1066092 CGAGCAGGAGTACTGACATCTGCACATCTCA-----GCCCATTTCAACTCTTG 1066145
Qy 91 AsnTyrArgGluIleGlyHisIleAsnAsnLeuAsnSerSerTyrThrIleTyrThrGly 110
Db 1066146 ATGTGG-----CATGCACAAAAC-----GGC 1066166
Qy 111 GlnThrLeuThrLeuThrSerGlyAspLeuLysValArgGluArgSerIleSerSerGly 130
Db 1066167 CAATGATTCGCAAAAGTTCTTATCTTATCTTATCTTGGACTCAATTCATGACATTCGCAAT 1066226
Qy 131 ValAsnThrAlaHisThrProSerProValAlaValGlnSerSerArgProProValGln 150
Db 1066227 ATCAATACTGA-----ACGCAAGCTTTGTAGTACTTGAACGTCCTTTAAAC 1066277
Qy 151 GlnHisProAla-----ValGlnLysProThrProProValValValValLys 166
Db 1066278 GGAACACCTGAAACGTAGCACTTGGGCACAAAGAGTGCTTAACAAATTTGTTAACTTAA 1066337
Qy 167 LysProThrProThrProProValValGlnGlnProAlaProValAlaProProValThr 186
Db 1066338 ATCCCAACCGGTGGCGAGGTTATATTT-----GCTCCAAATTTCAAGTCTTATTC 1066388
Qy 187 GluLysProPheAlaThrGlySerSerGlyValMetGlnPheArgTyrProValGlyAla 206
Db 1066389 -----GTRACAGCAAGCAATGGGT----- 1066406
Qy 207 ThrAsnProValValArgArgPheGlyThrAlaThrValAlaGlySerThrValThrSer 226
Db 1066407 -----TGAGAACGAGATCCATCCCGGAGCGCAGAAATTTTAC 1066445
Qy 227 AsnGlyMetTyrPheSerGlyArgAspGlyAspLeuIleAsn----- 240
Db 1066446 AATGCTATG-----GACTTGTATTAATGGCAATCCAAACAATCCA 1066484
Qy 241 -----AlaSerAsnAlaGlyThrValIleGlnIleAlaAspHisAsn-----MetAsp 255
Db 1066485 ATCTTAGCTTTCTGGCGATGCTCAAGTGTCCAGCGGAGGAATATATATGACTGGAT 1066544
Qy 256 GlyAlaSerIleValIleGlnHisThrAsnGlyPheValSerSerTyrIleHisIleLys 275
Db 1066545 GGAATATTACCGGTCATCAACATGCGGATGATGATTTATACAGGGTACGCATCAAGC 1066604
Qy 276 AspAlaGlnValLysThrGlyAspThrValArgThrGlyGlnArgIleAlaSerMet 294
Db 1066605 AGAATTCATGTTCTGTGGTCAAAATGTATAAAAGGCAACAATGAGCTTATG 1066661

```

```

RESULT 33
AI199683/c
ID AI199683 standard; DNA; 4403765 BP.
AC AI199683;
XX
DT 15-JAN-2002 (first entry)

```

```

XX DE Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 2.
XX KM Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;
XX KM variation; epidemiology; patient treatment; epidemic monitoring; db.
XX OS Mycobacterium tuberculosis.
XX SS US6294328-B1.
XX PM 25-SEP-2001.
XX PF 24-JUN-1998; 98US-0103840.
XX PR 24-JUN-1998; 98US-0103840.
XX PA (GENO-) INST GENOMIC RES.
XX PI Fleischmann RD, White OR, Fraser CM, Venter JC;
XX DR WPI; 2001-647261/74.
XX PT Evaluating strain variation of Mycobacterium tuberculosis, comprises
XX PT determining the nucleotide sequence of the strain at positions in the
XX PT genome corresponding to positions where M. tuberculosis strains CDC
XX PT 1551 and H37Rv differ.
XX PS Claim 4; SEQ ID NO 2; 3bp + Sequence Listing; English.
XX CC The invention relates to evaluating strain variation within and between
XX CC different populations of the tuberculosis bacterial pathogen.
XX CC Mycobacterium tuberculosis or related Mycobacterium by determining the
XX CC nucleotide sequence of the first strain at positions in the complete
XX CC sequence of the genome that correspond to positions that differ in the
XX CC nucleotide sequences of M. tuberculosis strains CDC 1551 (AI199683) and
XX CC H37Rv (AI199682). The method is useful for evaluating strain variation of
XX CC M. tuberculosis and has valuable application in the fields of
XX CC tuberculosis genetics, epidemiology, patient treatment and epidemic
XX CC monitoring.
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from USPRO
XX CC at seqdata.uspto.gov/sequence.html?docID=6294328B1.
XX SQ Sequence 4403765 BP; 757105 A; 1447799 C; 1441301 G; 757371 T; 189 other;

Alignment Scores:
Pred. No.: 1.46e+04 Length: 4403765
Score: 130.50 Matches: 81
Percent Similarity: 34.37% Conservative: 41
Best Local Similarity: 22.82% Mismatches: 128
Query Match: 7.81% Indels: 106
DB: 22 Gaps: 17

US-10-018-706-2 (1-322) x AI199683 (1-4403765)
Qy 40 AsnSerThrSerGlySerGlySerHisArgThrSerGlySerGlyLeuAlaIleGly 59
Db 431307 AACTCGGCGACCGGCAACTGCGGTATCGGCACCTCGGCACCGGCAACCGGCATTCGGC 431248
Qy 60 SerGlnValIleThrAsp-----SerGlnGlyValProAsnArgTyrGlnVal 75
Db 431247 AACACCGGACGACATTAACACGGGCTTTCAACACCGGCAATCGTCACACCGGCTGCC 431188
Qy 76 LysGlnGlyAspThrValSerLysIleAlaGlnArgTyrGlyLysLeuAsnTyrArgGluIle 95
Db 431187 AACCGGCGACG-----TACAAACCGGCTGTGTACAAACAC 431152
Qy 96 GlyHisIleLeu----- 99
Db 431151 GCGCACACCAACACCGGCGATGCCAACCCTGGGCACTTCAACACGGGCTTCAACACCC 431092
Qy 100 ---AsnLeuAsnSerSerTyrThr-----IleTyrThrGlyGlnThrLeuThr 114

```

```

DB 431091 GGCATTTCAGTACGGGCTTTGGCCAAACGAGGTGATATGCGCACCGGGGCTTTCATCACC 431092
QY 115 -----LeuTPserglyAspLeuIyValArgIleuArgSerIle 127
DB 431031 GGGGACATGGGGAACGGGCGCTTCTGGCGGGGAC-----CAGACGGCCATATC 430981
QY 128 SerSerGlyValAsnThr-----AlaHisThrProSerProValAla 141
DB 430980 AGCGCGGCTATCGGGTCCATGTTCCGGAATACCCGACAGCTCACCGTGAAGTTCCC 430921
QY 142 ValGln-SerSerArgProProValGlnGlnHisProAlaValGlnIyProThrPro 161
DB 430920 GTCAACATCCCATCCATCAGCCGACCT-----TCACCAACACCGTTC 430882
QY 161 oValValVal-----ValIyIyProThrProThrProProValValGlnGlnProAl 179
DB 430881 TACAGCGCATACGCTTAGCAMAATCACTTCGTTTACCATCCATCCGACATCCGAGGATC 430822
QY 179 aProValAlaProProValThrGlnAlaProPhaIalThrGlySerSerGlyValMetG 199
DB 430821 CCCCTGCTGGCCGTCGCAATCAAGCAAGCCGTTCTCCGCCATCACCGGACCGGTCCC 430762
QY 199 nPheArgTyrProValGlyAlaThr-----AsnProValValArgArgPheGlyTh 216
DB 430761 GCGATCACGCTCAACATCCGCGACCTCGCGGCTTCAGACCGCG-ATCAGATCCCGGCCAC 430703
QY 216 rAlaThrValAlaGlySerThrValThr----- 225
DB 430702 CCGAAGCGTCGCTTCGATGTCACGTTGTCACATTCGGCGGTACACGAGGCTTTT 430643
QY 226 -----SerAsnGlyMetTrpPheSerGlyArgAspGlyAspLeuI 239
DB 430642 CAACGCCACCAACGATCCGTCCTCTCGGGC---TTCTTCAACGCGCGCCCGAACCTATC 430586
QY 239 eAsnAlaSerAsnAlaGlyThrValIleGlnAlaAspHisAsnMetAspGlyValaserI 259
DB 430585 GGGCATCGGCACATCGGCGGCCAACATTCCTCGGCTTCGACAGTGCACACTCGCG-- 430528
QY 259 eValIleGlnHisThrAsnGlyPheValaserGlyTyrIleHisIleIyAspAlaGlnVa 279
DB 430527 -----ACCTCGGGCTTC---AACAACTACGCGCTCGTTCAAATCGGGACTGCG 430484
QY 279 IlysthrGlyAspThrVal-----ArgThrGly-----GlnArgT 291
DB 430483 GAACCTGGGCGATACCGTCTCGGCGGTATTCAAACCGGCATCGGGGACCGCCAACT 430424
QY 291 eAlaSerMetIyAsn-----GlnProSerG 300
DB 430423 CTCGGGCATGTTCAACATCGGACGACAACTCGCGGGGTTCTTCACAGCACGAGCGACGG 430364
QY 300 yAlaAlaIleuPheGlnPheArgIleSerArgAsnGlyValTyr 314
DB 430363 GATGTCGATGTTCAACCTCGGCGCTGGGAACATCGGCCAATTC 430321

```

RESULT 34

```

AA199682/C
ID AA199682 standard; DNA; 4411529 BP.
XX
XX AA199682;
XX
XX 15-JAN-2002 (first entry)
XX
XX Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 1.
XX
XX Mycobacterium tuberculosis, strain H37Rv, strain CDC 1551, genome;
XX variation; epidemiology; patient treatment; epidemic monitoring; ds.
XX
XX Mycobacterium tuberculosis.
XX
XX US6294328-B1.
XX
XX 25-SEP-2001.
XX

```

```

PF 24-JUN-1998; 98US-0103840.
XX
XX 24-JUN-1998; 98US-0103840.
XX
XX (GENO-) INST GENOMIC RES.
XX
XX Fleischmann RD, White OR, Fraser CM, Venter JC;
XX
XX WPI; 2001-647261/74.
XX
XX
XX Evaluating strain variation of Mycobacterium tuberculosis, comprises
XX determining the nucleotide sequence of the strain at positions in the
XX genome corresponding to positions where M. tuberculosis strains CDC
XX 1551 and H37Rv differ
XX
XX
XX Claim 3; SEQ ID NO 1; 3bp + Sequence Listing; English.
XX
XX
XX The invention relates to evaluating strain variation within and between
XX different populations of the tuberculosis bacterial pathogen.
XX Mycobacterium tuberculosis or related Mycobacterium by determining the
XX nucleotide sequence of the first strain at positions in the complete
XX sequence of the genome that correspond to positions that differ in the
XX nucleotide sequences of M. tuberculosis strains CDC 1551 (AA199683) and
XX H37Rv (AA199682). The method is useful for evaluating strain variation of
XX M. tuberculosis and has valuable application in the fields of
XX tuberculosis genetics, epidemiology, patient treatment and epidemic
XX monitoring.
XX
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from USPTO
XX at segdata.uspto.gov/sequence.html?docID=6294328B1.
XX
XX
XX Sequence 4411529 BP; 758565 A; 1449983 C; 144602 G; 758379 T; 0 other;
XX
XX
XX Alignment Scores:
XX Pred. No.: 1.46e+04 Length: 4411529
XX Score: 130.50 Matches: 81
XX Percent Similarity: 34.37% Conservative: 41
XX Best Local Similarity: 22.82% Mismatches: 128
XX Query Match: 7.81% Indels: 106
XX DB: Gaps: 17
XX
XX
XX US-10-018-706-2 (1-322) x AA199682 (1-4411529)
XX
XX
XX 40 AsnSerThrSerGlySerGlySerHisArgThrSerGlySerGlyIleuAlaIleGly 59
DB 431224 AACTCGGGCACCGGCACACTGGGATCGGCACACTCCGCGACGACACCGGCATCGGC 431165
QY 60 SerGlnValIleThrAsp-----SerGlnGlyValProAsnArgTyrGlnVal 75
DB 431164 AACACCGGCACGACATACACGCGGCTTCTTCAACACCGGCATCGTCACACCGGTGCGC 431105
QY 76 LysGlnGlyAspThrValSerIyIleAlaGlnArgTyrGlyLeuAsnTrpArgGlnIle 95
DB 431104 AACCGGGCGACG-----TACAAACCGGCTGGTATACACAC 431069
QY 96 GlyHisIleAsn----- 99
DB 431068 GCGACACCAACACCGCATCGGCACACTCGGCGCATTCACACGAGCTTATACACAC 431009
QY 100 ---AsnLeuAsnSerSerTyrThr-----IleTyrThrGlyGlnTrpLeuThr 114
DB 431008 GGCATTTTCATGACGGGCTTTCACCAACGAGGTATATCGGCACCGGCTTTCATACCC 430949
QY 115 -----LeuTPserGlyAspLeuIyValArgIleuArgSerIle 127
DB 430948 GCGCATGGGCAACGCGCTTCTGCGCGGCAC-----CAGCAGGCTATATC 430898
QY 128 SerSerGlyValAsnThr-----AlaHisThrProSerProValAla 141
DB 430897 AGCGCGGCTATCGGGTCCATGTTCCGGAATACCCGACAGCTCACCGTGAAGTTCCC 430838
QY 142 ValGln-SerSerArgProProValGlnGlnHisProAlaValGlnIyProThrPro 161

```

Db	430837	GTCAACATCCCATCACCAGCT	-----TCAACACACGTC	430799
Qy	161	oValValVal	-----VallyslYProThrProThrProProValValInglInProAl	179
Db	430798	TACACGGCATACGCTTAGCAAAATCAACTTCGGTTTCAACATCGCAGGATC	430739	
Qy	179	aProValAlaPProProValThrGlnuLaProPheAlaThnGlySerSerGlyValMetG	199	
Db	430738	CCCCCTGCTGGCGGTGCAATCGACAAAGCCGTTCTCCGCCCATCACCGGACCGGTCCC	430679	
Qy	199	nPheArgTyrProValGlyAlaThr	-----AsnProValValArgArpPheGlyTh	216
Db	430678	GGCATCACGGTCAACATCGCGCACCTCGGCGGTGACCGCG--ATCAGATTCGCCGGCAC	430620	
Qy	216	rAlaThrValAlaGlySerThrValThr	-----	225
Db	430619	CGCAAGCGTCGCTCCCTTGATGTACAGTTCGTCACATTCGCGGTACACGCGGCTTTT	430560	
Qy	226	-----SerAsnGlyMetTrpPheSerGlyValArgAspGlyAspLeuI	239	
Db	430559	CAACGCCACACCGATCCGCTCGTCCGGG--TTCCTTCAACGGCGGCCCGGAAACCGTATC	430503	
Qy	239	eAsnAlaSerAsnAlaGlyThrValIleGlnAlaAspHisAsnMetAspGlyAlaSerI	259	
Db	430502	GGGCAATCGCAACATCGGCGCCCAACATTCGCGCTTCCAGAACGTGCGCAATCCGG--	430445	
Qy	259	eValIleGlnIsthrAsnGlyPheValSerSerTyrIleHisIleLysAspAlaGlnVa	279	
Db	430444	-----ACCTGGGGCTTC--AACAACTAACGGTGGCTGCATCGGAGACTGGC	430401	
Qy	279	IllysthrGlyAspThrVal	-----ArgThrGly	291
Db	430400	GAACCTGGCGCATACCGTCTCGGCGGTATTCAAACCGGATCGGGGACCGGCCAACGT	430341	
Qy	291	eAlaSerMetLysAsn	-----GlnProSerG	300
Db	430340	CTCGGGCATGTTCACACATCGGACGACCACTCGCGGGGTTCTTCCACGACCGAGCGACCGG	430281	
Qy	300	YAlaAlaLeuPheGlnPheArgIleSerArgAsnGlyValTyr	314	
Db	430280	GATGTGCATGTTCACACTCGGCTCGGGAACATCGGCCCAATTC	430238	
RESULT 35				
AAL46625				
ID	AAL46625	standard; DNA; 1233 BP.		
AC	AAL46625;			
XX				
DT	05-AUG-2002	(First entry)		
XX				
DE	H influenzae BASB201 coding sequence #2.			
XX				
KW	BASB201; otitis media; pneumonia; sinusitis; nosocomial infection;			
KW	auditive nerve damage; delayed speech learning; vaccine;			
KW	antibacterial; auditory; antiinflammatory; gene; ds.			
XX				
OS	Haemophilus influenzae.			
XX				
FT	Key	Location/Qualifiers		
FT	CDS	1..1233		
FT		/*tag= a		
FT		/product= "BASB201"		
XX				
PN	WO200230967-A2.			
XX				
PD	18-APR-2002.			
XX				
PF	05-OCT-2001; 2001WO-EP11561.			
XX				
PR	13-OCT-2000; 2000GB-0025169.			
XX				
PA	(SMIK) SMITHKLINE BEECHAM BIOLOGICALS.			
XX				

PI Thomard J;
DR WPI: 2002-426267/45.
XX P-PSDB; AAO17657.
XX
XX New isolated non-typable Haemophilus influenzae BASB201 polypeptides,
PT useful as components of vaccines for treating bacterial infection such
PT as otitis media, delayed speech learning and inflammation of middle ear
PT
XX
XX Claim 13; Page 87-88; 90pp; English.

CC The present invention provides the protein and coding sequences of
CC several versions of the BASB201 protein from non-typable Haemophilus
CC influenzae. These can be used in the production of vaccines against H.
CC influenzae infection, which can cause otitis media in infants and
CC children, pneumonia in elders, sinusitis, nosocomial infections, or
CC invasive diseases, chronic otitis media with hearing loss, fluid
CC accumulation in the middle ear, auditive nerve damage, delayed speech
CC learning, infections of the upper respiratory tract and inflammation of
CC the middle ear. The present sequence is a version of the BASB201 coding
CC sequence of the invention.

SQ Sequence 1233 BP; 463 A; 226 C; 241 G; 303 T; 0 other;

Alignment Scores:
Pred. No.: Length: 1233
Score: 130.00 Matches: 64
Percent Similarity: 38.19% Conservative: 33
Best Local Similarity: 25.20% Mismatches: 111
Query Match: 7.78% Indels: 46
DB: 24 Gaps: 11

US-10-018-706-2 (1-322) x AAL46625 (1-1233)

Gy 74 GlnVallylGInglyAspThrValSerIleAlaGlnArgTylGlyLeuAsnTrpArg 93
Db 562 CAAAAAAAACAACAAACAAGATTGTCAAAAAGAACAAGCAGAACATCAA----- 609
Gy 94 GluileGlyHisIleAsnAsnLeuAsnSerSeryrThrlleTythrGlyIntPleu 113
Db 610 -----TCACCGCTGAATGAATCAATAAATAAATTAGCCCTTGATCAAGATAA---- 657
Gy 114 ThrLeutrpserGlyAspLeuylsValArgIuaRserIleserSerglyValAsnThr 133
Db 658 -----TTGAATGCCCTTAATAAGCAACAAACAAGCATTCGTCAAGAAATTCAACA 708
Gy 134 AlaHistrProserProvalAlaValGinSerSeraRgProPro-----ValGln 150
Db 709 GCTGAA-----CAGCAGATGGCGCAACAAGAAAAAGTGAAGAAGGCATTCGTCA 762
Gy 151 GlnHisProAlaValGlnlyseProThrProProvalValVallylsybProthrPro 170
Db 763 CGCCAAAAAGCGAAGAAAAAGAACAA----- 789
Gy 171 ThrProProvalValGlnlmpProalProValAlaProProValthrGluAlaProphe 190
Db 790 -----TCAAAACCTTATCAACAACACTGTGCAAGAACCCCAATT 828
Gy 191 AlaThrGlyserSerglyVal-----MetGlnPheArgTyrProValGlyAla 206
Db 829 CTTAATAGTACAAAGCGGTTTAGGGGGCGCAAAAAAACAATATTCCTTACAGTTTCGT 888
Gy 207 ThrAsnProvalValArgArpPhelYthrAlatrValAlaGlySerThrValThrSer 226
Db 889 TCA-----ATTTCGCATATCTTTTGT-----TCATATCAAGCGCAAGATCGTTGG 936
Gy 227 AsnglyMetTrpPheserGlyArgAsrglyAspLeuIleAsnAlaserAsnlaaglyThr 246
Db 937 AAAGGTATAGTATATTGGCGCATCAGACAGCACCGCTGTAAAGCAATTGCTGTGACGC 996
Gy 247 ValileGlnAlaAspHis---AsmdeAsrglyAlaserIleValilleGlnIstrAsn 265

Db	997	GTCAATTTAGCCGGCATATTAAATGGTTATNGTTATATGTTATGTTAAACACGGCGAA	1056
Oy	266	GlypHeValserSerYrTleHsIleLyAspAlaGlnValIysThrGlyAspThrVal	285
Db	1057	ACTGATTAAAGTTATTAATGAGCTTCATCATCAAGCGGTATCAAGTGAAGTGTGCGCTTT	1116
Oy	286	ArgThrGlyGlnArgIleAlaIasermetLysAsn-----GlnProserGlyAlaIaleu	303
Db	1117	TCACAGAGGGCAGGTATTATGCTCAGTGAATACACAGGGAATATACAGTTCTGGCGTT	1176
Oy	304	PheGluPheArgIleSerArgAsnGlyValIyrValAspPro	317
Db	1177	TAT---TTGGTATTAAGCCGTAAAGAAACGCCAGCAATAATCCT	1215
RESULT 36			
ID	AAZ49721	standard, DNA, 1520 BP.	
XX	AAZ49721;		
XX	18-APR-2000	(first entry)	
XX		Achromobacter lyticus beta-lytic protease gene.	
XX		Beta-lytic protease; antibiotic; anti-staphylococcal;	
XX		gene therapy; mastitis; staphylococcal infection; ruminant; cow;	
XX		transgenic animal; altered beta-lytic protease gene; ds.	
XX		Achromobacter lyticus.	
XX	Key	Location/Qualifiers	
XX	CDS	319..1443	
XX		/*tag= "a	
XX		/product= "Beta-lytic protease"	
XX	MO9967381-A1.		
XX	29-DEC-1999.		
XX	22-JUN-1999;	99WO-US14073.	
XX	22-JUN-1998;	98US-0090175.	
XX	21-JUN-1999;	99US-0337079.	
XX	(UYVE-) UNIV VERMONT & STATE AGRIC COLLEGE.		
XX	Bramley JA, Plaut KI, Kerr D;		
XX	WPI, 2000-147208/13.		
XX	P-PSDB; AAY44646.		
XX		Treatment of Staphylococcal infections, such as mastitis, in ruminant	
XX		animals, especially cows	
XX		Disclosure; Fig 14A; 61pp; English.	
XX		The present sequence is a Achromobacter lyticus DNA encoding	
XX		beta-lytic protease which has anti-staphylococcal activity.	
XX		The present sequence is used in the production of altered genes which	
XX		allow expression and preferably secretion of active protein in mammalian	
XX		cells/tissues. The altered gene is produced by operably linking the	
XX		beta-lytic protease coding sequence with mammalian promoter, signal	
XX		peptide and translation initiation sequences. The modified sequence	
XX		is used in gene therapy to treat staphylococcal mastitis infections in	
XX		ruminants, e.g. goats, sheep, and cows. It is also used to produce	
XX		transgenic animals which are resistant to staphylococcal infections.	
XX	Sequence 1520 BP; 258 A; 530 C; 500 G; 232 T; 0 other;		
Alignment Scores:			
Pred. No.:	1.36	Length:	1520
Score:	128.50	Matches:	79
Percent Similarity:	33.98%	Conservative:	43

Best Local Similarity:	22.01%	Mismatches:	118
Query Match:	7.69%	Indels:	119
DB:	21	Gaps:	15

US-10-018-706-2 (1-322) x AAZ49721 (1-1520)

QY	10	GlhansGlnLysProIleLysArgLeuGlyLeuIlePheGlyValIleThrThrCysIle	29
DB	313	CAGCATATGAAAGATTTCGAAGCGGAGCTG-----GGGTGGCGCTGGTGTCCGG	366
QY	30	LeuAlaGlyCysAlaSerLysProThrTyAsnSerThrSerGlySerGlySerHisArg	49
DB	367	CTGGCG-----ACGATCGCGGCAACGACGCGCGAG	399
QY	50	ThrSerGlySerGlyGlyLeuAlaIleGlySerGlnValIleThrAspSerGlnGlyVal	69
DB	400	GGCAGCGGCTCAG-----CGCGAGATCTGGTGTATTCTACAGACGAGATG-----	444
QY	70	ProHisArgTyGlnValIleGlnGlyAspThrValSerLysIleAlaGlnArgTyGly	89
DB	445	-----TTGCATCTGCATCATCATCGCATCTGGCGCAAGCATGCGCGCATCTGCAC	495
QY	90	LeuAsnTrpArgGluIleGlyHisIleAsnLeuAsnLeuAsnSerTyThrIleTyThr	109
DB	496	AAACCTTCGGAAAGATCTCGC-----	519
QY	110	GlyGlnTrpLeuThrLeuTrpSerGlyAspLeuLysValArgGluArgSerIleSerSer	129
DB	520	-----TGGCGCGC-----TACAGCGGATCAGCGCA	546
QY	130	GlyValAlaSerThrAlaHisThrProSerProValAlaValGlnSerSerArgProProVal	149
DB	547	AGGTTCATCGCGCTGATGAGCAGCAGACCGCGCGCTCACCGCAAGCGCGCAGCATC	606
QY	150	GlnGlnHisProAlaValAlaGlnLysProThrProProValValValLysLysPro	168
DB	607	GTTCGTTCCGGCAAGCTGGCGCGCGCGCAGC-----GCTTCGGCGCGCAACCCGCG	657
QY	169	-----ThrProThrProProValValGln-----	176
DB	658	AGGTGCGCTGCGCTGCGCGAGTTCGCTTACAGCCGATCCCGACGCGCCAAAGGGGCC	717
QY	176	-----	176
DB	718	GGTGAAGCTGGCGCGCGCAATCCGCTGCAAGCGCTTTCAGCGCTTCGGCGCAACA	777
QY	177	-----GlnProAlaProValAlaProProValThr	186
DB	778	GCCGGCGCGCGCTGCGCGCGGAGCGGAGTTCAGCTGTGCTTACGCGCGCTGTTC	837
QY	187	Glu--AlaPro-----Phe	190
DB	838	CGAACCCTGCGCGAGCCAGCGCGCTTCGACCGCTTCGCAAGCGCGCGCGCGAGCTGCA	897
QY	191	AlaThrGlySerSerGlyValMetGlnPheArgTyProValGlyAlaThrAsnProVal	210
DB	898	GCCGTGTCGCCAACCGCGCTGTGCAAGTTCCTTCGCGCGCGCGCCACGCTGGATGTC	957
QY	211	ValArgArgPheGlyThrAlaThrValAlaGlySer-----	222
DB	958	-----GGCGGCGCCACCAACCAACACCGCGCTCGGCAATTACCGCATGCTGTCGTC	1008
QY	223	ThrValThrSerAsnGlyMetTrpPheSerGlyArgAspGlyAspLeuIleAsnAlaSer	242
DB	1009	GACATGTGCGCGCGCGCGCGCGCGCGCGCGACCAACGCAACGCAACTGCGTGTGCGCTCG	1068
QY	243	AsnAlaGlyThrValIleGlnAlaAspHisAsnMetAspGlyAlaSerIleValIleGln	262
DB	1069	GGCGCGCGCTGTTCACACGC-----CACTTCCTGCTTCCTTCGCGGAGATGTG-----	1116
QY	263	HisThrAsnGlyPheValSerSerTyThrIleHisIleLysAspAlaGlnValLysThrGly	282
DB	1117	CACACCGCGCGCTGTGTCACACACTTACACACTGATGAAACATTCGATACACACCGCG	1176

Qy 283 AsphThValArgThrGlyGlnArgIleAlaSerMetLysAsnGlnProSerGlyAla 301
 Db 1177 GCCAACGTCGAGGAAACACCGCCATCCCAACCGCGCAACACCCAGCGCGCG 1233

RESULT 37
 AAX20250 ID AAX20250 standard; DNA; 111309 BP.
 AC AAX20250;
 XX
 XX 04-MAY-1999 (first entry)
 XX
 XX Borrelia burgdorferi polynucleotide sequence #3.
 DE
 XX Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;
 KM epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;
 KM infection; diagnosis; characterisation; detection; ds.
 XX
 OS Borrelia burgdorferi.
 XX
 PN MO9858943-A1.
 XX
 PD 30-DEC-1998.
 XX
 PF 18-JUN-1998; 98WO-US12764.
 XX
 PR 03-SEP-1997; 97US-0057483.
 PR 20-JUN-1997; 97US-0050359.
 PR 22-JUL-1997; 97US-0053344.
 PR 22-JUL-1997; 97US-0053377.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 PA (MEDI-) MEDIMUNE INC.
 XX
 PI Clayton R, Dougherty BA, Fraser C, Iathigra R, Smith HO;
 PI White OR;
 DR WPI; 1999-081217/07.
 XX
 XX New isolated Borrelia burgdorferi nucleic acids - used to develop
 PT products for the detection, diagnosis, characterisation, prevention
 PT and therapy of infections, particularly Lyme disease
 XX
 XX Claim 1; Page 738-800; 1128pp; English.
 PS
 XX AAX20248 to AAX20402 represent polynucleotide sequences isolated from
 CC Borrelia burgdorferi (Bb). Products derived from Bb can be used for
 CC the detection, diagnosis, characterisation, prevention and therapy of
 CC Bb infections, e.g. Lyme disease. They can also be used for the
 CC production of biosynthetic products, e.g. enzymes. Borrelia belongs
 CC to a family of motile, spiral-shaped bacteria called Spirochetes.
 CC Spirochetes are pathogenic in humans and Borrelia causes epidemic and
 CC endemic relapsing fever, and Lyme borreliosis, more commonly known as
 CC Lyme disease.
 CC
 XX Sequence 111309 BP; 35956 A; 13151 C; 19075 G; 43117 T; 10 other;
 SQ

Alignment Scores:
 Pred. No.: 243 Length: 111309
 Score: 128.50 Matches: 64
 Percent Similarity: 33.71% Conservative: 25
 Best Local Similarity: 24.24% Mismatches: 94
 Query Match: 7.69% Indels: 81
 DB: 20 Gaps: 9

US-10-018-706-2 (1-322) x AAX20250 (1-111309)

Qy 69 ValProAsnArg-----TyrGlnValLysGlnGlyAspThrValSerLysIle 84
 Db 45502 GTTCCCAATATGAAAGCAATTCTTTATATTTGTTAAAAAATGACTCTATCTCATCTATA 45561

Qy 85 AlaGlnArgTyrGlyLeuAsnTrpArgGlnIleGlyHisIleAsnLeuAsnSerSer 104

Db 45562 GCTAGTCTTATATGATTCCTCCCAAGTTGATATTTAGATCTATATATCTGATATGAA 45621
 Qy 105 TyrThrIleTyrThrGlyGlnTrpLeuThrLeuTrpSerGlyAspLeuValArgGlu 124
 Db 45622 --GTTTATTTTATTTGAGCAAAAGTTGTTATCTCTGGGGAGAAATG-- 45666
 Qy 125 ArgSerIleSerSerGlyValAsnThrAlaHisThrProSerProValAlaValGlnSer 144
 Db 45666 ----- 45666
 Qy 145 SerArgProProValGlnGlnHisProAlaValGlnLysProThrProProValVal 164
 Db 45667 -----CCCAAGAATTTTAAAGAG 45687
 Qy 165 ValLysLysProThr-----ProThrProProValValGlnGlnProAlaProVal 181
 Db 45688 GTATTAGGGAGACTTTTATTTATTCCTGCGAGGCTTTATTTACTTCGGGGTATGGCTAT 45747
 Qy 182 AlaProProValThrGlnAlaProPheAlaThrGlySerSerGlyValMetGlnPheArg 201
 Db 45748 CGACCAAGT-----CCGTTT-----ACAGAGTTATTAGTTTTCAC 45783
 Qy 202 TyrProValGlyAlaThrAsnProValValArgArgPheGlyThrAlaThrValAlaGly 221
 Db 45783 ----- 45783
 Qy 222 SerThrValThrSerAsnGlyMetTrpPheSerGlyValArgAspGlyAspLeuIleAsnAla 241
 Db 45784 -----AATGGAATGATATTCGCAAAATTTAGCTAATACGCCCTATTAAGCC 45828
 Qy 242 SerAsnAlaGlyThrValIleGlnAlaAspHisAsnMetAspGlyAlaSer-----Ile 259
 Db 45829 TCAAGAGAAGGTGTTGTTGTTACTGACGATTTAATCGGAGGCTATGCAAAATATATTT 45888
 Qy 260 ValIleGlnHisThrAsnGlyPheValSerSerTyrIleHisIleLysAspAlaGlnVal 279
 Db 45889 GTCATTTCTCAGACCAACGAGATTCCAAACTTATATGACATTTGAATTCCTTTGGCGTT 45948
 Qy 280 LysThrGlyAspThrValArgThrGlyGlnArgIleAlaSerMetLysAsnGlnPro--- 298
 Db 45949 AAGGTTGAAAAAAGTTTCAAGGGGCGGTATATAGTTATATGGAAGCACTGCGCTAT 46008
 Qy 299 SerGlyValAlaLeuPheGlnPheArgIleSerArgAsnGlyValTyrValAspProLeu 318
 Db 46009 AGTACGGGCAATCATTTTGCATTTTACCATTTTAAATGTAAGTAAGCAAAATCTATG 46068
 Qy 319 ThrValLeuLys 322
 Db 46069 AATATTTTAAAGA 46080

RESULT 38
 AAX20248 ID AAX20248 standard; DNA; 910715 BP.
 AC AAX20248;
 XX
 XX 04-MAY-1999 (first entry)
 XX
 XX Borrelia burgdorferi polynucleotide sequence #1.
 DE
 XX Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;
 KM epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;
 KM infection; diagnosis; characterisation; detection; ds.
 XX
 OS Borrelia burgdorferi.
 XX
 PN MO9858943-A1.
 XX
 PD 30-DEC-1998.
 XX
 PF 18-JUN-1998; 98WO-US12764.
 XX

PR 03-SEP-1997; 97US-0057483.
 PR 20-JUN-1997; 97US-0050359.
 PR 22-JUL-1997; 97US-0053344.
 PR 22-JUL-1997; 97US-0053377.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 PA (MED1-) MEDIMUNE INC.
 XX
 PI Clayton R, Dougherty BA, Fraser C, Lathigra R, Smith HO;
 PI White OR;
 DR WPI; 1999-081217/07.
 PT New isolated *Borrelia burgdorferi* nucleic acids - used to develop
 PT products for the detection, diagnosis, characterisation, prevention
 PT and therapy of infections, particularly Lyme disease
 PS Claim 1; Page 157-671; 1128pp; English.
 XX
 CC AAX20248 to AAX20402 represent polynucleotide sequences isolated from
 CC *Borrelia burgdorferi* (Bb). Products derived from Bb can be used for
 CC the detection, diagnosis, characterisation, prevention and therapy of
 CC Bb infections, e.g. Lyme disease. They can also be used for the
 CC production of biosynthetic products, e.g. enzymes. *Borrelia* belongs
 CC to a family of motile, spiral-shaped bacteria called Spirochetes.
 CC Spirochetes are pathogenic in humans and *Borrelia* causes epidemic and
 CC endemic relapsing fever, and Lyme borreliosis, more commonly known as
 CC Lyme disease.
 XX
 SQ Sequence 910715 BP; 327171 A; 129646 C; 130753 G; 323091 T; 54 other:
 Alignment Scores:
 Pred. No.: 3, 07e+03 Length: 910715
 Score: 128.50 Matches: 64
 Percent Similarity: 33.71% Conservative: 25
 Best Local Similarity: 24.24% Mismatches: 94
 Query Match: 7.69% Indels: 81
 DB: Gaps: 9
 US-10-018-706-2 (1-322) x AAX20248 (1-910715)
 QY 69 ValProAsnArg-----TyrGlnValIysGlnGlyAspThrValSerLysIle 84
 DB 634888 GTTCCCAATATGAAAGGAATGTTTATATGTTAAATAAAGACCTCATCTATATA 634947
 QY 85 AlaGlnArgTyrGlyLeuAsnTrpArgIuileGlyHisIleAsnAsnLeuAsnSerSer 104
 DB 634948 GGTAGTGTCTATATATGTTCCCAAGTGTGATTTATAGATTCTATATATCTGATATGAA 635007
 QY 105 TyrThrIleTyrThrGlyGlnTrpLeuThrLeuTrpSerGlyAspLeuLysValArgIu 124
 DB 635008 ---GTTTATTTTATGAGCAAAAGTTGTTATTCCTGGGGAAGATTG----- 635052
 QY 125 ArgSerIleSerSerGlyValAsnThrAlaHisThrProSerProValAlaValGlnSer 144
 DB 635052 ----- 635052
 QY 145 SerArgProProValGlnGlnHisProAlaValGlnLysProThrProProValValVal 164
 DB 635053 -----CCCAAGATTTTATAAAGAG 635073
 QY 165 ValIlyLysProThr-----ProThrProProValValGlnGlnProAlaProVal 181
 DB 635074 GATTATGGGAGACCTTTATTTATCTGTGACGGGTATTAATCTCGGGGTATGCTAT 635133
 QY 182 AlaProProValThrGlnAlaProPheAlaThrGlySerSerGlyValMetGlnPheArg 201
 DB 635134 CGACCAAGT-----CCGTTT-----ACGAGGATTAATGTTTTCAC 635169
 QY 202 TyrProValGlyAlaThrAsnProValValArgArgPheGlyThrAlaThrValAlaGly 221
 DB 635169 ----- 635169

QY 222 SerThrValThrSerAsnGlyMetTrpPheSerGlyArgAspGlyAspLeuIleAsnAla 241
 DB 635170 -----AATGGATATAGATATGCAATTTAGCAATATACCCCTTTAAAGCC 635214
 QY 242 SerAsnAlaGlyThrValIleGlnIleAspHisAsnMetAspGlyAlaSer-----Ile 259
 DB 635215 TCAAGAGAAGAGTGTGTTGTTATCTGACAGATTTATGCGGAGGATGAAATATATAT 635274
 QY 260 ValIleGlnHisThrAsnGlyPheValSerSerTyrIleHisIleLysAspAlaGlnVal 279
 DB 635275 GTCAATTTCTCACAGACGAGATCCCACTTATATGCAATTTGAATTCCTTTGCCGTT 635334
 QY 280 LysThrGlyAspThrValArgThrGlyGlnArgIleAlaSerMetLysAsnGlnPro--- 298
 DB 635335 AAGTTGGAAAAAAGTTTCAAGGGAGCGGTATATGTTATATGGAAGCACTGGCTAT 635394
 QY 299 SerGlyAlaAlaLeuPheGluPheArgIleSerArgAsnGlyValTyrValAspProLeu 318
 DB 635395 AGTACGGGCAATCATTTGCATTTTACATTTTAAATAATGTAAGAAATCTATATG 635454
 QY 319 ThrValLeuLys 322
 DB 635455 AATATTTAAGA 635466
 RESULT 39
 AAQ25083
 ID AAQ25083 standard; DNA; 1520 BP.
 AC AAQ25083;
 DT 15-NOV-1992 (first entry)
 XX
 DE Beta-lytic protease gene.
 XX
 KW Gram-negative; Gram-positive; bacteria; decomposition; ss.
 XX
 OS Achromobacter lyticus.
 XX
 FH Key Location/Qualifiers
 FT CDS 319..1443
 FT /tag=a
 XX
 PD JP04108387-A.
 XX
 PD 09-APR-1992.
 XX
 PF 29-AUG-1990; 90JP-0225136.
 XX
 PF 29-AUG-1990; 90JP-0225136.
 XX
 PR (WAKO) WAKO PURE CHEM IND.
 PA
 XX
 DR WPI; 1992-171653/21.
 DB P-PSDB; AAR24147.
 XX
 PT Beta-lytic protease gene and DNA encoding it - for decomposing
 PT Gram-positive and some Gram-negative bacteria
 XX
 PS Claim 1; Fig 1; 13pp; Japanese.
 XX
 CC Peptidase was purified from *Achromobacter lyticus* by a Sepharose
 CC CL-4B column, a Sephadex G-75 column, and by reverse phase HPLC.
 CC The first 25 N-terminal amino acids were found to be identical to
 CC those of the beta-lytic protease from *Lysobacter enzymogenes*. The
 CC beta-lytic protease gene of *A. lyticus* was cloned using PCR primers
 CC based on the sequence of the *L. enzymogenes* beta-protease. The
 CC full length sequence of the gene encoding the *A. lyticus* enzyme was
 CC determined by diideoxy sequencing. The beta-lytic protease is
 CC expected to be an enzyme which can decompose not only Gram-positive
 CC bacteria but also some Gram-negative bacteria.
 CC See also AAQ25084-6.
 XX
 SQ Sequence 1520 BP; 258 A; 533 C; 498 G; 231 T; 0 other;

Alignment Scores:

Pred. No.: 1.62 Length: 1520
 Score: 127.50 Matches: 79
 Percent Similarity: 33.70% Conservative: 42
 Best Local Similarity: 22.01% Mismatches: 119
 Query Match: 7.63% Indels: 119
 Gaps: 15

US-10-018-706-2 (1-322) x AA025083 (1-1520)

```

Qy 10 Glnaanglnlysprolllelyasrgleuylleuilephegilyvalillethrtthrcysalle 29
Db 111 Glnaanglnlysprolllelyasrgleuylleuilephegilyvalillethrtthrcysalle 29
Db 313 CAGACCATGAGAGATTCCAGAGCGGAGCTG-----GGGCTGCGCTGCTGCTGCGCG 366
Qy 30 leuilaaglycybalaaserlyspthrtyraanserthrseryglyserthiaryg 49
Db 367 CTGGCG-----ACGATCGCGCGGCGCAAGCGCGCGCGCG 399
Qy 50 Thrseryglyserglyleuallailegysergvalillethraspsrglnlyval 69
Db 400 GCCACGCTCAG-----CGGAGGATCTGCTGATTTCTACAGACGAGATG----- 444
Qy 70 Proanaarytyrglnvallysglnlyaspthryalserlyleallaiglnarytygly 89
Db 445 -----TTGCACTTCGACATCGATCGCATTCGCGCAAGCATGCCGCGATCTGCAC 495
Qy 90 leuantaarytyrglnlyleaglyhialleuanaanserlythrtthrtthrtthr 109
Db 496 AAGCACTCGAAGAGATCTCGCAC----- 519
Qy 110 Glylntrpleuhtleuhtleuhtleuhtleuhtleuhtleuhtleuhtleuhtleuht 129
Db 520 -----TGGGCGCG-----TACAGCGGATCGACCGCA 546
Qy 130 Glyvalaanthrthraahsthrproserprovalalalalinserseraryproproval 149
Db 547 AGTGTGATCGCGCTGATGAGACAGACAGCGCGCGGTCACGCCAAGCGCGCGACGATC 606
Qy 150 Glnlnlnhsprollalavalglnlyspthrproprovalalvalalvallylspbro-- 168
Db 607 GTCGCTTCGCGAGCATCGCGCGCGCGCGCGCG-----GCTTCGCGCGCGCGCGCG 657
Qy 169 -----ThrProthrProprovalalalaln----- 176
Db 658 AGTTCGCGCTGCGCGCTGCGCGAGTCGCTGTACAGCGCGATCCGACCGCGCGCGCGCG 717
Qy 176 ----- 176
Db 718 GGTGACGCTGCGCGCGCGCGCATTCGCGCTGTCGAGCGCTTCGCGCGCGCGCGCAACGA 777
Qy 177 -----GlnProalaprovalalalaproprovalthr 186
Db 778 GCCGCGCGCGCGCTGCGCGCGCGCGAGTTCAGTGTCTACAGCGCGCGCTGTTCAA 837
Qy 187 Glu---Alapro-----Phe 190
Db 838 CGAACCGCGCGCGCGCGCGCGCGCTTCGCGCGCGCTTCGCCAAGCGCGCGCGCGACGTGCA 897
Qy 191 AlaThrglyserserglyvalalmetglnphearytyrprovalalalathraspProval 210
Db 898 GCCGTCGCGCGCGCGCGCGCTGCGAGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 957
Qy 211 ValarytyrphleglythralaThralaalyser----- 222
Db 958 -----GGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1008
Qy 223 ThrValThrseryamgylmettrpPheSerglyaragapglayaspLeuileasnaaser 242
Db 1009 GACATGTCGCGCGCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1068
Qy 243 Asnaiaaglythralalleglnalaspnaaserlyasnaaserlyalaserlevalilegln 262
  
```

```

Db 1069 GCCCGCGCGCTGTCGAAGCGC-----CACTCTTCGTCGCTTCGCGAGATGCTG----- 1116
Qy 263 HistrhaanglyphevalaserSerythlethlethlyasapalaglnvallythrgly 282
Db 1117 CACACCGCGCGCTGTCGACACCTGATCAACCTGATGACATCACTGATCAACACCGCGC 1176
Qy 283 AspThryalagthrglyglnaryglnalalasermetlysaanglnproserglyala 301
Db 1177 GCCAAGTGTGATGAGAACCGCGCATCGCCACCGCGCGCGCGCGCGCGCGCGCGCG 1233

RESULT 40
AAZ00266
ID AAZ00266 standard; DNA; 7277 BP.
XX
AC AAZ00266;
XX
DT 12-OCT-1999 (first entry)
XX
DE HEV-US2 full length gene sequence.
XX
KW Hepatitis E virus; HEV; binding partner; virus; US-HEV infection;
KW vaccine; passive immunisation; ss.
XX
OS Hepatitis E virus.
XX
PN MO9919732-A1.
XX
PD 22-APR-1999.
XX
PF 15-OCT-1998; 98WO-US21941.
XX
PR 15-OCT-1997; 97US-0061199.
XX
PA (ABBO ) ABBOTT LAB.
PI Dawson GJ, Desai SM, Erker JC, Mushahwar IK, Schlauder GG;
DR WPI; 1999-288017/24.
XX
PT Detection of United States isolates of hepatitis E virus
XX
PS Claim 36; Page 201-204; 260pp; English.
XX
CC The invention provides a method for detecting a US (sub)type hepatitis E
CC virus (US-HEV), or its naturally occurring variants in a sample by
CC treatment with a binding partner specific for a marker of the virus, and
CC then detecting any complex formed. The method is used to diagnose
CC infection with US-HEV. Polypeptides from US-HEV, antibodies specific for
CC open reading frames (ORF) in US-HEV and host cells expressing these ORFs
CC are useful in vaccines or for passive immunisation. The polypeptides are
CC also used to raise specific antibodies (useful as immunoassay reagents).
CC Fragments of nucleic acid from US-HEV are useful as primers and probes in
CC usual hybridisation and amplification assays for detecting infection. The
CC present sequence represents a HEV-US2 full length gene sequence.
XX
SQ Sequence 7277 BP; 1333 A; 2111 C; 1904 G; 1916 T; 13 other;

Alignment Scores:
Pred. No.: 11.7 Length: 7277
Score: 127.00 Matches: 80
Percent Similarity: 33.14% Conservative: 37
Best Local Similarity: 22.66% Mismatches: 120
Query Match: 7.60% Indels: 116
Gaps: 19

US-10-018-706-2 (1-322) x AAZ00266 (1-7277)

Qy 27 Thrcysilleuilaaglycybalaaserlyspthrtyraanserthrsery 43
Db 1755 ACGAGGAGTGTGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1814
Qy 44 Glyserlyserthrasgthrseryglyserglyleuallaile----- 58
  
```

```
Db 1815 TTGACGCTCTGCGAGTCTATGAGGGCGGGGTGCGACAGCCTCACTTATGAGCTCACC 1874
Qy 59 -----GlySerGlnValIleThrAspSerGlnGlyValProAsnArgTyrglnValLys 76
Db 1875 CCGGCCGGCTGTCAGGTAAGATTTCATCTAATGCTGTGATTCAGTGCACCTGCCACATTCGCC 1934
Qy 77 GlnGlyAspThrValSer----- 82
Db 1935 CCYGGTGGGCGCCCTGAGCCCGCGGGGAGGTGGCGSCTTCGCAGTGCTCTTAT 1994
Qy 83 -----LysIleAlaGlnArgTyrglyLeuAsn-----TrpArg----- 93
Db 1995 AGATACATATAGTTCACCCAGCGGCATTCGCTGACGCGGAGCTATGGCTACATCCTGAG 2054
Qy 94 -----GluIleGlyHisIle-----AsnAsn 100
Db 2055 GGGCTGCTGGTATCTTCCCCCATTTCTCCCTGGGCATATTTGGAGTCTGCTAACCCC 2114
Qy 101 LeuAsnSerSerTyrrThrlIeTyrrGlyGlnTrpLeuThrIeu---TrpSerGlyAsp 119
Db 2115 TTTTGGCGTGAAGGGGACTTGTATATACCCGACCTGGTCAACCTTGTTTCTAGTGAAT 2174
Qy 120 LeuLysValArgGluArgSerIleSerSerGlyValAsnThrAlaHisThrProSerPro 139
Db 2175 TTCTCCCCCTGAGCGCGCTCTCTGCTGCTGCGCTGCCGCCGCCGGGTTGGCCTACCT 2234
Qy 140 ValAlaValAlaGlnSerSerArgProProVal----- 149
Db 2235 ACT-----CCACCTGTATGATATCTGGGTGTATACACGCCCTCA 2276
Qy 150 ---GlnGlnHis-----ProAlaValGlnLysPro-----Thr 159
Db 2277 GAGGAATCTCATGTGATGAGCGGATCTGATACCTCTGTTCCGAGCTGCTGATTAACC 2336
Qy 160 ProProValValValValLysLysProThrProThrProProValValGlnGlnProAla 179
Db 2337 AGCCCTATGTGCTTAC-----CCCCCCCCCTCTCCCTGCGTAAAGCCGACA 2390
Qy 180 ProValAlaProProValThrGluAlaProPheAlaThrGlySerSerGlyValMetGln 199
Db 2391 ACATCCCGCCTCCCGCAGCTGCCGT-----CTCCTT 2423
Qy 200 PheArgTyrrProValGlyAlaThrAsnProValValArgArgPheGlyThrAlaThrVal 219
Db 2424 TACACCTACCCCGACGCGGCC-----AAGGTGTAAT 2453
Qy 220 AlaGlySerThrValThrSerAsnGlyMetTrpPheSerGlyArgAspGlyAspLeuIle 239
Db 2454 GCGGGGTCTATGTGTGAGTCAGACATGCTGATTTG-----TTAGTTC 2492
Qy 240 AsnAlaSerAsnAlaGlyThrValIleGlnAlaAspHisAsnMetAspGlyAlaSerIle 259
Db 2493 AATGCTCAAAACCCCTGGC-----CATCGCCCGGGGGTGGC----- 2528
Qy 260 ValIleGlnHisThrAsnGlyPheValSerSerTyrrIleHisIleLysAspAlaGlnVal 279
Db 2529 ---CTCTGCCAT-----GCTTTTATCAACGTTTCCAGAAAGCGTTTAC 2570
Qy 280 LysThrGlyAspThrValArgThrGlyGlnArgIleAlaSerMetLysAsnGlnPro--- 298
Db 2571 TCGACTGAATTCATCATGCGGAGGCGCTTGACGATACACTTAAACCCCGGCCCTAT 2630
Qy 299 SerGlyAlaAlaLeuPheGluPheArgIleSerArgAsn 311
Db 2631 ATCCATGCACTGGCTCCGACTATAGGTTGAGCAAAAC 2669
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_p2n model

Run on: July 3, 2003, 20:12:12 ; Search time 48 Seconds
(without alignments)
2057.290 Million cell updates/sec

Title: US-10-018-706-2
Perfect score: 1670
Sequence: 1 MVTAINSONQKRIKILG.....LFEFRISRNGVYDPLTVLK 322

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 441362 seqs, 15338381 residues
Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DBV=x1h
-Q=/cgn2_1/USFTO.spool/US10018706/runat_30062003_091105_23816/app_query.fasta_1.519
-DB=Issued Patents NA -OPMT=fastap -SUFFIX=nti -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=40
-MODE=LOCAL -OUTFMT=pct -NOEM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10018706 @CNC 1.1 32 @runat_30062003_091105_23816 -NCPU=6 -ICPU=3
-NO MAP -LARGEROBJ -NEG_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DET_TIMEOUT=120 -MAIN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents NA:*
1: /cgn2_6/prodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/prodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/prodata/1/ina/6A.COMB.seq:*
4: /cgn2_6/prodata/1/ina/6B.COMB.seq:*
5: /cgn2_6/prodata/1/ina/PCUS.COMB.seq:*
6: /cgn2_6/prodata/1/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	219.5	13.1	1215	5 PCT-US96-05320A-707	Sequence 707, Appl
2	209	12.5	2179	4 US-09-405-728-4	Sequence 4, Appl1
3	165	9.9	1885	3 US-08-619-812-3	Sequence 3, Appl1
4	131	7.8	3646	3 US-08-619-812-7	Sequence 7, Appl1
5	130.5	7.8	1930	2 US-08-737-716-1	Sequence 1, Appl1
6	130.5	7.8	1930	2 US-08-737-716-11	Sequence 11, Appl1
7	130.5	7.8	15872	4 US-09-105-537-1	Sequence 1, Appl1
8	130.5	7.8	4403765	4 US-09-103-840A-2	Sequence 2, Appl1
9	130.5	7.8	4411529	4 US-09-103-840A-1	Sequence 1, Appl1
10	123.5	7.4	4411529	4 US-09-103-840A-1	Sequence 1, Appl1
11	120.5	7.2	3004	1 US-08-276-213-6	Sequence 6, Appl1
12	119.5	7.2	11707	4 US-09-136-574A-1	Sequence 1, Appl1

13	118.5	7.1	4403765	4 US-09-103-840A-2	Sequence 2, Appl1
14	118	7.1	5127	4 US-09-462-606-1	Sequence 1, Appl1
15	118	7.1	7207	4 US-09-462-606-3	Sequence 3, Appl1
16	117.5	7.0	2830	2 US-09-010-928B-1	Sequence 1, Appl1
17	116	6.9	4518	3 US-09-125-287-2	Sequence 2, Appl1
18	116	6.9	12839	3 US-09-125-287-1	Sequence 1, Appl1
19	115	6.9	3300	1 US-08-194-290-6	Sequence 6, Appl1
20	115	6.9	3300	2 US-08-614-377A-6	Sequence 6, Appl1
21	115	6.9	3300	4 US-09-142-648B-6	Sequence 6, Appl1
22	113.5	6.8	1896	4 US-09-311-626B-15	Sequence 15, Appl1
23	113.5	6.8	7100	4 US-09-308-375-1	Sequence 1, Appl1
24	112.5	6.7	2186	2 US-08-878-546-9	Sequence 9, Appl1
25	112.5	6.7	8438	1 US-07-945-283-1	Sequence 1, Appl1
26	112	6.7	36519	3 US-08-923-137-2	Sequence 2, Appl1
27	111	6.6	867	4 US-09-071-035-921	Sequence 221, App
28	111	6.6	993	4 US-09-134-001C-1165	Sequence 1165, Ap
29	110.5	6.6	2077	1 US-08-217-327-7	Sequence 7, Appl1
30	110.5	6.6	6416	4 US-09-136-574A-2	Sequence 2, Appl1
31	109.5	6.6	2029	4 US-09-136-574A-46	Sequence 46, Appl1
32	109	6.5	3833	1 US-08-917-320-18	Sequence 18, Appl1
33	109	6.5	3833	5 PCT-US95-04611A-18	Sequence 18, Appl1
34	109	6.5	5931	3 US-08-783-774-1	Sequence 1, Appl1
35	109	6.5	5931	4 US-09-556-706B-1	Sequence 1, Appl1
36	108	6.5	3172	1 US-08-314-309A-1	Sequence 1, Appl1
37	107.5	6.4	2061	2 US-09-258-371-9	Sequence 9, Appl1
38	107.5	6.4	2061	3 US-08-751-230-9	Sequence 9, Appl1
39	107.5	6.4	2061	3 US-09-489-082-9	Sequence 9, Appl1
40	107.5	6.4	2061	4 US-09-258-372-9	Sequence 9, Appl1
41	107.5	6.4	2061	4 US-09-006-783A-2	Sequence 2, Appl1
42	107	6.4	1470	4 US-09-109-841-1	Sequence 1, Appl1
43	106.5	6.4	1446	4 US-09-149-476-139	Sequence 139, App
44	106.5	6.4	8367	2 US-08-583-562B-7	Sequence 7, Appl1
45	106.5	6.4	8367	2 US-08-779-113-7	Sequence 7, Appl1

ALIGNMENTS

RESULT 1
PCT-US96-05320A-707
Sequence 707, Application PC/TUS9605320A
GENERAL INFORMATION:
APPLICANT: Human Genome Sciences
APPLICANT: 9410 Key West Avenue
APPLICANT: Rockville, MD 20850
APPLICANT: United States of America
APPLICANT: Johns Hopkins University
APPLICANT: 720 Rutland Avenue
APPLICANT: Baltimore, MD 21205
APPLICANT: United States of America
APPLICANT: Mark D. Adams
APPLICANT: Owen White
APPLICANT: Hamilton O. Smith
APPLICANT: J. Craig Venter
TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus Influenzae Rd Genome,
NUMBER OF SEQUENCES: 48
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox
STREET: 1100 New York Avenue, Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20003-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/05320A
FILING DATE: April 22, 1996
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/476,102

FILING DATE: June 7, 1995
 PRIORITY DATE: June 7, 1995
 APPLICATION NUMBER: 08/487,429
 FILING DATE: June 7, 1995
 ATTORNEY/AGENT INFORMATION:
 NAME: Eric K. Steffe
 REGISTRATION NUMBER: 36,688
 REFERENCE/DOCKET NUMBER: 1488.014PC01
 TELEPHONE: (202) 371-2540
 TELEFAX: (202) 371-2540
 INFORMATION FOR SEQ ID NO: 707:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1215 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 PCT-US96-05320A-707

Alignment Scores:
 Pred. No.: 5,08e-12 Length: 1215
 Score: 219.50 Matches: 76
 Percent Similarity: 44.69% Conservative: 46
 Best Local Similarity: 27.84% Mismatches: 114
 Query Match: 13.14% Indels: 37
 DB: Gaps: 10

US-10-018-706-2 (1-322) x PCT-US96-05320A-707 (1-1215)

71 AaAaGTYGlnValIysGlnIyAaPThrValIserIyIleAglAaGlyAaGlyLeu 90
 433 AATACCTTAATTAAGTAAACAAAGCGATCAATGTTCTTAATGCTTACTGCGCGGAT 492
 91 AaAaGTYGlnValIysGlnIyAaPThrValIserIyIleAglAaGlyAaGlyLeu 110
 493 GATGTAAAGAAATTCGCGCGCTTGAATTAACCTATCCGAACCTTAATTAAGTTAGG 552
 111 GlnTPLeuThrLeuTrpSerGlyAaPLeuValAaGlyAaGlySerIleSerGly 130
 553 CAAGTTTAAAGATTTCTAATGTCATATAAAGGTTACTACAAAGGTTCT 606
 131 ValAaThrAlaIleThrProSerProValAlaValAaGlnSerSerArgProProValAla 150
 607 -----GTAACAACTCGACGACACAGCATCAACGCAACGCAAGTTAA 651
 151 GlnHisProAlaValAaGlnIySerProThrPro----- 160
 652 -----CAAGCATCTTAT---ACTCCGGCGCAACCGTACTCAATATGATCTGAC 702
 161 -----ProValValValValIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIyIy 175
 703 GGTACTATTATTCGCCCAATTAATCAAGAGCGTGCACATCGCTTAACTGTCGCA 762
 176 GlnGlnProAlaProValAlaProProValAlaThrglnAlaProPheAlaThrglySer 195
 763 ACTTCATCAACACAGGTTACATCTTAAATATATGCAATAGCAGCAACCAATTAATTC 822
 196 GlyValMetGlnPheArgTyProValGlyAlaThraPheProValAlaArgPheGly 215
 823 AATGCTGTTGGC-----CCAAATC---GCTTCAAT---GTTGTGTCGACATGCGCG 867
 216 ThrAla-----ThrValAlaGlySerThrValThr-----SerAaGlyMetTrp 230
 868 ACTTCAGGTATATATCAATCAAGGTTCTTCAAGCAGATGCGCGTAAACAAAGGAATGAT 927
 231 PheSerGlyAaGlyAaPLeuIleAaPLeuIleAaPLeuIleAaPLeuIleAaPLeuIle 250
 928 ATTAGTGGTTACACAGGCAAGCTGTAAAGCGGTCGACAGGCGCAATGTCGTATGCA 987
 251 AaPheIleMetAaPLeuIleAaPLeuIleAaPLeuIleAaPLeuIleAaPLeuIle 268
 988 GGCATAGCTTACGTTGCTTACGTTAATTTATTTATTTATTTATTTATTTATTTATTT 1047

269 SerSerTyIleIleHisIleLeuAaPLeuValAaGlnValIySerThrglnAlaPheThrgly 288
 1048 AGTGCCTTGGCATTAACGAAATTTCTTCCCGCATCAACAGAAAGTCAAGCAGCT 1107
 289 GlnArgIleAaSerMetIyAaGlnIyProSerGlyAlaAlaLeuPheArgIle 308
 1108 CAAGCATCGCAAAATGGTACGCTGTCGTAATACCTGAAATCTTCACCTTGAAAT 1167
 309 SerArgGlyValIyValAaPLeuThrValIleu 321
 1168 CGCTATAAGTAAATCATCATGATGATCATGATGATGATGATGATGATGATGATGAT 1206

US-09-405-728-4
 Sequence 4, Application US/09405728
 Patent No. 6391316

GENERAL INFORMATION:
 APPLICANT: Potter, Andrew A.

APPLICANT: Schryvers, Anthony B.

TITLE OF INVENTION: CLONING AND EXPRESSION OF HAEMOPHILUS SOMNUS

FILE REFERENCE: 9000-0049.20

CURRENT FILING DATE: 1999-09-24

EARLIER FILING DATE: 1999-03-10

NUMBER OF SEQ ID NOS: 5

SOFTWARE: Patent In Ver. 2.0

SEQ ID NO 4

LENGTH: 2179

TYPE: DNA

ORGANISM: Haemophilus somnus

FEATURE:

NAME/KEY: CDS

LOCATION: (872)..(1906)

US-09-405-728-4

Alignment Scores:
 Pred. No.: 1.3e-10 Length: 2179
 Score: 209.00 Matches: 63
 Percent Similarity: 44.09% Conservative: 49
 Best Local Similarity: 24.80% Mismatches: 106
 Query Match: 12.51% Indels: 36
 DB: Gaps: 6

US-10-018-706-2 (1-322) x US-09-405-728-4 (1-2179)

73 TyGlnValIyGlnIyAaPThrValIserIyIleAglAaGlyAaGlyLeuAaPThr 92
 1229 TACAAAGTACGCAAAAGCGATCAATGTTCTTAATGCTTACTGCGCGGATTA 1288
 93 ArgGlnIleGlyHisIleAaPLeuAaPLeuAaPLeuAaPLeuAaPLeuAaPLeu 112
 1289 AAGAAATGCGCACATTAATATATGTCGACCATCATCATCATCATCATCATCAT 1348
 113 LeuThrLeuTrpSerGly-----AaPLeuValAaGlyAaGlySerIleSer 129
 1349 TTGAAATGCAAAATATATTCCTCCGATGCAATATGATACCAACAGCAATTAATGA 1408
 130 GlyValAaPLeuIleHisThrProSerProValAlaValAaGlnSerSerArgProProVal 149
 1409 TCGAGGTACACAAATATCATGTCATGAG--ACATGAAAGCTATTAATTAACCAAT 1465
 150 GlnGlnHisProAlaValAaGlnIyProThrProProValAlaValAaGlyIyProThr 169
 1466 GAAACAAATGAACCCGTTGCTACACACATTCACCAATGCAATCAATCAATCAAT 1519
 170 ProThrProProValAlaGlnIyProAlaProValAlaProProValAlaThrglnAlaPro 189
 1520 ---ACACCTTCAGCCCACTCAATATATATGATGATGATGATGATGATGATGAT 1558
 190 PheAlaThrglySerSerGlyValMetGlnPheArgTyProValGlyAlaThraPhePro 209

Db 1559 -----ACAAATGGAAAA 1570
Qy 210 ValValArgArgPheGlyThrAlaThrValAlaGlySerThrValThrsAsnGlyMet 229
Db 1571 ATTATTCAGAGATTTCCAGTCTGATGAGGC-----AATTAAGGTATT 1615
Qy 230 TrpPheSerGlyArgAspGlyAspLeuIleAsnAlaSerAsnAlaGlyThrValIleGln 249
Db 1616 GATATTACGGTTCCTCGGACCAAGCTGTATTATGACGACGCTGGACGAGTTGTATAT 1675
Qy 250 AlaAspHisAsnMetAspGlyAlaSer-----IleValIleGlnHisThrAsnGlyPhe 267
Db 1676 GCCCGAGACGCTTACGTCGATATGCTATTATTAATTATTAACATTAATGACAGCTTAT 1735
Qy 268 ValSerSerGlyrIleHisIleLeuAspAlaGlnValIleGlyThrGlyAspThrValArgThr 287
Db 1736 TTAAGTCTTATGACATTAAGATGATGCTGTAAGATCAGCAAGAGATTAAACG 1795
Qy 288 GlyGlnArgIleAlaSerMetLeuAsnGlnProSerGlyAlaAlaLeuPheGlyPheArg 307
Db 1796 GGTCAACAATTTCTTAATATGGAAGTCTTGAAACAACAATCAACTCATTTTGAA 1855
Qy 308 IleSerArgAsnGlyValIleValAspProLeuThrValIleu 321
Db 1856 ATTCGTATAAGTCAATCAATGATGATCAATGAGATATTTA 1897
RESULT 3
; Sequence 3, Application US/08619812
; Patent No. 6100066
; GENERAL INFORMATION:
; APPLICANT: POTTER, ANDREW A.
; APPLICANT: THEISEN, MICHAEL
; APPLICANT: HARLAND, RICHARD J.
; TITLE OF INVENTION: VACCINES FOR HAEMOPHILUS SOMNUS
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: REED & ROBIN
; STREET: 635 BRYANT STREET
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/619,812
; FILING DATE: 15-MAR-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/038,719
; FILING DATE: 29-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: ROBINS, ROBERTA L.
; REGISTRATION NUMBER: 33,208
; REFERENCE/DOCKET NUMBER: 9000-0019.20
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 617-8999
; TELEFAX: (415) 327-3231
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1885 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURES:
; NAME/KEY: CDS

LOCATION: 872..1708
US-08-619-812-3
Alignment Scores:
Pred. No.: 1.85e-06 Length: 1885
Score: 165.00 Matches: 57
Percent Similarity: 43.80% Conservative: 49
Best Local Similarity: 23.55% Mismatches: 98
Query Match: 9.88% Indels: 38
DB: 3 Gaps: 6
US-10-018-706-2 (1-322) x US-08-619-812-3 (1-1885)
Qy 73 TyrGlnValIleGlySerGlyAspThrValSerIleAlaGlnArgGlyLeuAsnTrp 92
Db 1229 TACAAGATGCCAAGGAGATACATGTTCTTATGCTTATTTCCATGCGATGATATA 1288
Qy 93 ArgGluIleGlyHisIleAsnAsnLeuAsnSerSerGlyrIleThrIleGlyGlnTrp 112
Db 1289 AAAGATTGGCCACACTTAATATATATGCTGAGCCATATCATCTGATGACAGTA 1348
Qy 113 LeuThrIleuTrpSerGly-----AspLeuYsValArgGluArgSerIleSerSer 129
Db 1349 TTGAATAATTCGAATTAATATATCCCATGACATATGATGATACACACACATTAATGA 1408
Qy 130 GlyValAsnThrAlaHisIleThrProSerProValAlaValGlnSerSerArgProProVal 149
Db 1409 TCAGAGGTGACACAAATACAGTCATAG--ACATGAAATGCTTAATTAACCAACAAT 1465
Qy 150 GlnGlnHisProAlaValGlnIleProThrProProValValValValIleYsProThr 169
Db 1466 GAACAATGGAACCGGTGCTACACCAACATTCACACATTCACACATCAATATAA----- 1519
Qy 170 ProThrProProValValGlnGlnProAlaProValAlaProProValThrGluAlaPro 189
Db 1520 --ACACCTCCAGCCACCTCAATATATGCTTGATTGGCCA----- 1558
Qy 190 PheAlaThrGlySerSerGlyValMetGlnPheArgTrpProValGlyAlaThrAsnPro 209
Db 1559 -----ACAAATGGAAAA 1570
Qy 210 ValValArgArgPheGlyThrAlaThrValAlaGlySerThrValThrsAsnGlyMet 229
Db 1571 ATTATTCAGAGATTTCCAGTCTGATGAGGC-----AATTAAGGTATT 1615
Qy 230 TrpPheSerGlyArgAspGlyAspLeuIleAsnAlaSerAsnAlaGlyThr--ValIleG 249
Db 1616 GATATTACGGTTCCTCGGACCAAGCTGTATTATGACGACGCTGACGACGACGAGTTGTAT 1675
Qy 249 IlaIaAspHisAsnMetAspGlyAlaSer-----IleValIleGlnHisThrAsnGlyP 267
Db 1676 ATCCCGGAGACGCTTACGTCGATATGCTATTATTAATTATTAATTAATGACAGCTT 1735
Qy 267 heValSerSerGlyrIleHisIleLeuAspAlaGlnValIleGlyThrGlyAspThrValArgT 287
Db 1736 ATTTAAGTCTTATGACATTAATGAAGTATACGCTGAAGATCAGCAAGAGATTAAAG 1795
Qy 287 hrGlyGlnArgIleAlaSerMetLeuAsnGlnProSerGlyAlaAlaLeuPheGlyPhe 307
Db 1796 CGGGTCAACAATAATGCTTAATATGGAAGTCTGGAACAACAACAATCAACTCATTTTTA 1855
Qy 307 rg 307
Db 1856 AA 1857
RESULT 4
US-08-619-812-7
; Sequence 7, Application US/08619812
; Patent No. 6100066
; GENERAL INFORMATION:
; APPLICANT: POTTER, ANDREW A.
; APPLICANT: THEISEN, MICHAEL
; APPLICANT: HARLAND, RICHARD J.

APPLICANT: RIOUX, CLEMENT R.
TITLE OF INVENTION: VACCINES FOR HAEMOPHILUS SOMNUS
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: REED & ROBINS
STREET: 635 BRYANT STREET
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 94301
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/619,812
FILING DATE: 15-MAR-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/038,719
FILING DATE: 29-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: ROBINS, ROBERTA L.
REGISTRATION NUMBER: 33,208
REFERENCE/DOCKET NUMBER: 9000-0019.20
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 617-8999
TELEFAX: (415) 327-3231
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 3646 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..2772
NAME/KEY: CDS
LOCATION: 2776..3570
US-08-619-812-7
Alignment Scores:
Pred. No.: 0.00986 Length: 3646
Score: 131.00 Matches: 48
Percent Similarity: 43.75% Conservative: 43
Best Local Similarity: 23.08% Mismatches: 79
Query Match: 7.84% Indels: 38
DB: 3 Gaps: 6
US-10-018-706-2 (1-322) x US-08-619-812-7 (1-3646)
QY 73 TYGIVNVALVSGINGLYASPTVRVLSERLYSILAGLNUAGTYRGLYLEUAENTP 92
DB 3091 TCAAGATGCGCAAGGCGATCCATGTTCTTATTCCTATATTCAGCGATGATTA 3150
QY 93 ATCGUULIEGLYHISTLEAENLEUAENSESETYTHRILEYRTHRGYGIHTNP 112
DB 3151 AAGAGATTGGCCACACTAATATATATGCTGAGCCATCACTGATTTGGACAGTA 3210
QY 113 LEUTHRIETPSEKLY-----ASPLEULYVALRGILURGSERIESESER 129
DB 3211 TTGAATAATGCAATATATTTCCCGATACCAATATGATACCAACAGCAATTAATGTA 3270
QY 130 GIVYALASHTHALAHISTHRPROSERPROVALALAVAGLINSERISERAGPROVAL 149
DB 3271 TCAGAGGTGACACCAATATCAGTCATAGG--ACATGGAGCTCATATTAACCAACAAAT 3327
QY 150 GINGINHISPRALAVAGLINSYPROTHPRPROVALVALVALVALYLSYSPROTHR 169
DB 3328 GAACAAATGAAACCGCTGTCTACACCAACACACATTCAACATGCAATCAATAA----- 3381

QY 170 PROTHRPROVALVALGINDINPROVALAIPROVALALAPROVALTHRGULALAPRO 189
DB 3382 ---ACACCTCGACGCCACCTCAATATATGCTTGATTTGGCCA----- 3420
QY 190 PHEALATHRGYSEKSERGLYALWECTGHPHEAGTYRPROVALGVALATHRANPPO 209
DB 3421 -----ACAAATGCAAAA 3432
QY 210 VALVALARGARPHGELYTHRALATHRVALALAGLYSETHRVALTHSERANGLYMET 229
DB 3433 ATTATTCAGAGATTTCTCGATCTATATGAGGC-----AATTAAGGTATT 3477
QY 230 TRIPHESEGLYARGASPOLYASPLEULEANALASERANALAGLYTHR--VALIIEG 249
DB 3478 GATATTAGCGGTTCTCGTGAGACAGCTGTTAATGACAGCGCTGACGCGAGTTGTAT 3537
QY 249 INALASPHISANMELASPOLYALASER-----ILEVALIEGLINHISTHRANGLYP 267
DB 3538 ATGCCGAGACCGCTTACGATATGATATTATTAATTATTAACATATATGACAGTT 3597
QY 267 HEVALSERSEYTHRIEHS 273
DB 3598 ATTTAAGCTTATGACAT 3617
RESULT 5
US-08-737-716-1
Sequence 1, Application US/08737716
Patent No. 5955258
GENERAL INFORMATION:
APPLICANT: GIBBE, BUIST
APPLICANT: GERARD VENEMA
APPLICANT: USN KOK
TITLE OF INVENTION: Process for the lysis of a culture of lactic
TITLE OF INVENTION: acid bacteria by means of a lysin, and uses of the resulting
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pillsbury Madison & Sutro, L.L.P.
STREET: 1100 New York Avenue, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005-3918
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Microsoft Word
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/737,716
FILING DATE: 22-APR-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/NL95/00170
FILING DATE: 12-MAY-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 94201353.3
FILING DATE: 12-MAY-1994
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1930 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Lactococcus lactis
STRAIN: MG1363
IMMEDIATE SOURCE:
CLONE: Fig.4 (publ.1)-acma and ORPA
FEATURE:

```

NAME/KEY: CDS
LOCATION: 178..1489
FEATURE:
NAME/KEY: CDS
LOCATION: complement (1542..1824)
US-08-737-716-1

Alignment Scores:
Score: 0.00412 Length: 1930
Percent Similarity: 130.50 Matches: 73
Best Local Similarity: 37.71% Conservative: 39
Query Match: 24.58% Mismatches: 96
Index: 89
Gaps: 14
DB: 2

US-10-018-706-2 (1-322) x US-08-737-716-1 (1-1930)

QY 26 ThrThrCysIleLeuAlaGlyCysAlaSerLysProThrTyraSerThr----- 42
Db 727 ACTGCTGCTTGAAGGAGATATGCTACCGACCAAGTATGGCGCTTCACTGAATCGC 786
QY 43 -----SerGlySerGlySerHisArgThrSergly 52
Db 787 ATTAATTCCTCAATATTAATTGACTCGTTTGAAGGAGCTTCTTCACTGAATATCTAAT 846
QY 53 SerGlyGlyLeuAlaIleGlySerGlnValIleThrAspSerglnGlyValProAsn--- 71
Db 847 TCTGTGTGC-----TCGACACACCAATTCAGATAATATTCGGAACCAATATGC 897
QY 72 -----ArgTyrGlnValIleGlnGlyAspThrValSerLysIleAlaGlnArgTyr 88
Db 898 AGTTCACACTACTATACCGCTCAATTCGTGATCTCTTGGGAATCTCAAAAGATAT 957
QY 89 GlyLeuAsnTrpArgGlnIleGlyHisIleAsnAsnLeuAsnSerglyThrIleTyr 108
Db 958 GGAATTAGTGTGCTCAATTCAAAGTCGAGATATCTTAAAGTATAC--ATTATCTAC 1014
QY 109 ThrGlyGlnTrpLeuThrLeuTrpSerglyAspLeuLysValArgGluArgSeriLys 128
Db 1015 ATTGCTCAAAACCTTGTACTACAGGTCA-----GCTTCTTCACAAAATTC 1062
QY 129 SerGlyValAsnThrAlaHisThrProSerProValAlaGlnSergSerArgProPro 148
Db 1063 GGTGCTTCAACCAATTC-----GCAGACACTACCTCAACC 1098
QY 149 ValGlnGlnHisProAlaValGlnLysProThrProProValValValLysLysPro 168
Db 1099 ACTTCTGTGACACCTGCA-----AAACCACTTCACAAACACTGTTAAGTTAAATCC 1152
QY 169 ThrProThr-----ProProValValGlnInProAlaProValAlaProProValThr 186
Db 1153 GGAATACCTTGGGGCGCTATCGATTAATAATAAATACTATTTGCTCAATGAAAAGT 1212
QY 187 GluAlaProPheAlaThrGlySerSerglyValMetGln-----PheArgTyrPro 203
Db 1213 TGAATCATTTAAGTTAGATACATTATATGCTGAATCTTATGTTTCAACATCT 1272
QY 204 ValGlyValAlaTrpAsnProValValArgArgPheGlyThrAlaTrpValAlaGlySerThr 223
Db 1273 GCTGCTGCTTCAAAATCC-----TCGACAGGTTGAGGCTCAACT 1311
QY 224 ValThrSerAsnGlyMetTrpPheSerglyArgAspGlyAspLeuIleAsnAlaSerAsn 243
Db 1312 GCTACCAATAC-----TCAAAC 1329
QY 244 AlaGlyThrValIleGlnAlaAspHisAsnMetAspGlyAlaSerIleValIleGlnHis 263
Db 1330 TCG----- 1332
QY 264 ThrAsnGlyPheValSerSeryTrpIleHisIleLysAspAlaGlnValLysThrGlyAsp 283
Db 1333 ACTTCTTCACTCAATGAGCTCAATTCAT-----AAGTGCTTAAAGAGAT 1380

```

```

QY 284 ThrVal-----ArgThrGlyGlnArgIleAlaSerMetLys 295
Db 1381 ACTGCTGGGAGACTTTCGCAAAATCTGGACGCCAATTCCTTCAATCAACAG 1431

RESULT 6
US-08-737-716-11
Sequence 11, Application US/08737716
Patent No. 5955258
GENERAL INFORMATION:
APPLICANT: Girbe BUIST
APPLICANT: Gerard VENEMA
APPLICANT: Jan KOK
APPLICANT: Adrianus Marinus LEDERBOER
TITLE OF INVENTION: Process for the lysis of a culture of lactic
TITLE OF INVENTION: acid bacteria by means of a lysin, and uses of the resulting
NUMBER OF INVENTION: lysed culture.
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESSES:
ADDRESSER: Pillsbury Madison & Sutro, L.L.P.
STREET: 1100 New York Avenue, N.W.
CITY: Washington
STATE: D.C. A.
COUNTRY: U.S.A.
ZIP: 20005-3918
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Microsoft Word
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/737, 716
FILING DATE: 22-APR-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/NL95/00170
FILING DATE: 12-MAY-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 94201353.3
FILING DATE: 12-MAY-1994
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 1930 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Lactococcus lactis
STRAIN: MG1363
IMMEDIATE SOURCE:
CLONE: Fig.4 - ORF8
FEATURE:
NAME/KEY: CDS
LOCATION: complement (1823..1928)
US-08-737-716-11

Alignment Scores:
Score: 0.00412 Length: 1930
Percent Similarity: 130.50 Matches: 73
Best Local Similarity: 37.71% Conservative: 39
Query Match: 24.58% Mismatches: 96
Index: 89
Gaps: 14
DB: 2

US-10-018-706-2 (1-322) x US-08-737-716-11 (1-1930)

QY 26 ThrThrCysIleLeuAlaGlyCysAlaSerLysProThrTyraSerThr----- 42
Db 727 ACTGCTGCTTGAAGGAGATATGCTACCGACCAAGTATGGCGCTTCACTGAATCGC 786
QY 43 -----SerGlySerGlySerHisArgThrSergly 52
Db 787 ATTAATTCCTCAATATTAATTGACTCGTTTGAAGGAGCTTCTTCACTGAATATCTAAT 846

```

```

QY 53 SerGlyGlyLeuAlaIleGlySerGlnValIleThrAspSerGlnGlyValProAm--- 71
DB 847 TGTGGTGC-----TCGACACCAACATTAACAAATTAATTCCTGGACCAATAC 897
QY 72 -----ArgTyrGlnValIleGlnGlyAspThrValSerIleAlaGlnArgTyr 88
DB 898 AGTTCACTACTTAATACCGTCAAATCTGTGATACCTTTGGGGAATCCACAAAGATAT 957
QY 89 GlyLeuAsnTTPArgGluIleGlyHisIleAsnAsnLeuAsnSerSerTyrThrIleTyr 108
DB 958 GGAATATAGTGTGCTCAAAATTCMAAGTCGAAATTAATCTTAAAGTACCC--ATTATCTCG 1014
QY 109 ThrGlyGlnTTPLeuThrLeuTTPSerGlyAspLeuIleValArgGluArgSerIleSer 128
DB 1015 ATTTGGTCAAAACTTGTACTGACAGGTCA-----GCTTCTTCACAAAATTC 1062
QY 129 SerGlyValAsnThrAlaHisThrProSerProValAlaValGlnSerSerArgProPro 148
DB 1063 GGTGGTTCAAAACATTC-----GCAAGCACTACTCCAGACC 1098
QY 149 ValGlnGlnHisProAlaValGlnIleProThrProProValValValIleValIlePro 168
DB 1099 ACTTGTGTGACACCTGCA-----AAACCACTTCACAAACAGCTTAAGTTAAATCC 1152
QY 169 ThrProThr-----ProProValValGlnGlnProAlaProValAlaProProValThr 186
DB 1153 GGAAGATACCTTTGGGCGCTATCAGTAAATATAAAGTAACTAGTCTCAATGGAAAGT 1212
QY 187 GluAlaProPheAlaThrGlySerSerGlyValMetGln-----PheArgTyrPro 203
DB 1213 TGGATCATTTATAGTTCAGATACCATTTATATGTTCAAAATCTTATGTTTCAACATCT 1272
QY 204 ValGlyAlaThrAsnProValValAlaArgArgPheGlyThrAlaThrValAlaGlySerThr 223
DB 1273 GCTGCTGTCAAAATCT-----TCGACAGCTTACAGGCTCAACT 1311
QY 224 ValThrSerAsnGlyMetTTPPheSerGlyArgAspGlyAspLeuIleAsnAlaSerAsn 243
DB 1312 GCTACCAATATAC-----TCMAAC 1329
QY 244 AlaGlyThrValIleGlnAlaAspHisAsnMetAspGlyAlaSerIleValIleGlnHis 263
DB 1330 TCG----- 1332
QY 264 ThrAsnGlyPheValSerSerTyrIleHisIleIleAspAlaGlnValIleThrGlyAsp 283
DB 1333 ACTTCTTAACTCAATATGCTCAATTCAT-----AAGTCTTAAGAGAGAT 1380
QY 284 ThrVal-----ArgThrGlyGlnArgIleAlaSerMetIle 295
DB 1381 ACTCTCTGGGGACTTCGCAAAATGTGGACGCCCAATGCTTCATCACTAAG 1431

```

RESULT 7
US-09-105-537-1/C

Sequence 1, Application US/09105537A

Patent No. 6255202

GENERAL INFORMATION:
APPLICANT: Sherman, D.H.

APPLICANT: Liu, H.

APPLICANT: Xue, Y.

APPLICANT: Zhao, L.

TITLE OF INVENTION: DNA encoding methymycin and pikromycin

FILE REFERENCE: 600,438US1

CURRENT APPLICATION NUMBER: US/09/105,537A

NUMBER OF SEQ ID NOS: 43

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 1

LENGTH: 15872

TYPE: DNA

ORGANISM: Streptomyces venezuelae

Alignment Scores:

Pred. No.:	Score:	Percent Similarity:	Best Local Similarity:	Query Match:	DB:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
4	0.108	130.50	38.75*	25.31*	4	15872	81	43	115	81	18

US-10-018-706-2 (1-322) x US-09-105-537-1 (1-15872)

```

QY 34 AlaSerIleProThrTyrAsnSerThrSerGlySerHisArgThrSerGlySer 53
DB 3195 GCGTCGCGCCC-----TCGGGACAGACCGCCGACAGGCTCTCGGGGGAG 3151
QY 54 -----GlyGlyLeuAlaIleGlySerGlnValIle 63
DB 3150 GTGACCGCGCGGGTAGCGGCGACGCGCATGCCAGATATCGCATGCGCTCGCGGTGTG 3091
QY 64 ThrAsp-----SerGlnGlyValProAsnArgTyrGlnValIleGlnIleAspThrVal 81
DB 3090 TCCGCGCGGGGTGCGGGGGGTATCCCTCGGCCATCCGATCTCGCGCGCGCGGTG 3031
QY 82 SerIle-----IleAlaGlnArgTyrGlyLeuAsnTTPArgGlu----- 94
DB 3030 ACCAGGTGCGCCAGGTGGCGCGCGCGCGCGCGCGCGCGCGGTGTGCAAGACAGTCCG 2971
QY 95 IleGlyHisIleAsnAsnLeuAsnSerSerTyrThrIleTyrThrGlyGlnTTPLeuThr 114
DB 2970 CTGGGCAAGCGCAGTCCGATGTGTGACAGACCGCTCCGACG----- 2926
QY 115 LeuTTPSerGlyAsp-----LeuIleValArgGluArgSer 126
DB 2925 -----TCGACGACATAGAGAGATCGAAGCCAGCTCTTGAAGTAAAGCGCATTCG 2872
QY 127 IleSerSerGly-----ValAsnThrAlaHisThrProSerPro 139
DB 2871 ACCGCGGTGCGGTCTGCTACTCGACGAGCGCGCGAGTGTGCGCGACGCTCGGTCTC 2812
QY 140 ValAlaValGlnSerSerArgProProValGlnGlnHisProAlaValAlaGlnIleProThr 159
DB 2811 GTCTGTGCGCGCGCGCGCGATCACCG-----CCACCGCGACCGCGCGCTCG 2764
QY 160 ProProValValValIleValIleValProThrProThrProProValValGlnIleProAla 179
DB 2763 CCC-----TCGCTTCTGCGCTGCGCGCGATCACCGCGCGCGG 2722
QY 180 ProValAlaProProValThrGluAlaProPheAlaThrGlySer----- 194
DB 2721 CCGGTGCGCGCGCGCGGTCCCGATCGCGCGCGCGCGCGCGCGCGGTGCGCGCGG 2662
QY 195 -----SerGlyValMetGlnPheArgTyrPro-----ValGlyAlaThrAsnProValVal 211
DB 2661 CCGTCAACCAAGTGGCTTGGCGCTGGAAGCGGTAGTGTGGCAGGGGACACCTGACCTG 2602
QY 212 ArgArgPheGlyThrAlaThrValAlaGlySer-----ThrValThr 225
DB 2601 CCGGTGCTCCGTGCGCGCGCGCTCGATCGACGCTGTGCGCGCGCGACGAAGAGGTGTG 2542
QY 226 Ser-----AsnGlyMetTTPPheSerGlyArgAspGlyAspLeuIleAsnAlaSerAsn 243
DB 2541 AGTGGCGGCGAGCGAGCTCGGCTCGCGCGCGCC-----TTGCGCAGG 2497
QY 244 AlaGlyThrValIleGlnAlaAsp-----HisAsnMetAspGlyAlaSerIleValIleGln 262
DB 2436 GCGGAGACCGCGGTGCGCGCTCTGTGTGCGGTACGAGGTCCGCGCATGCGCGAGAG 2437
QY 263 HisThrAsnGlyPheValSerSerTyrIleHisIleIleAsp-----AlaGlnValIleThr 281
DB 2436 ACCCGTGGGACCGAGCTCCAGAGAGGTGTGCGCGCGCATTCCTCCAGGAGGTGTGAG 2377
QY 282 GlyAspThrValArgThrGlyGlnArgIleAlaSerMetIleAsnGlnProSerGlyAla 301

```


Db 2376 GCGTCAGGAAGCGTACGCGCTCGCG-----ACCTGATCCACCCAGTACTCGGAGATCG 2323

RESULT 8

US-09-103-840A-2/C

Sequence 2, Application US/09103840A

Patent No. 6294328

GENERAL INFORMATION:

APPLICANT: FLEISCHMAN, Robert D.

APPLICANT: WHITE, Owen R.

APPLICANT: FRASER, Claire M.

APPLICANT: VENTER, John C.

TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM

TITLE OF INVENTION: TUBERCULOSIS

FILE REFERENCE: 24366-20007.00

CURRENT APPLICATION NUMBER: US/09/103,840A

CURRENT FILING DATE: 1998-06-24

NUMBER OF SEQ ID NOS: 2

SOFTWARE: Patent In Ver. 2.1

SEQ ID NO 2

LENGTH: 4403765

TYPE: DNA

ORGANISM: Mycobacterium tuberculosis

FEATURE:

OTHER INFORMATION: CDC 1551

OTHER INFORMATION: "n" bases at various positions throughout the sequence

OTHER INFORMATION: represent a, t, c or g

US-09-103-840A-2

Alignment Scores:

Pred. No.:	654	Length:	4403765
Score:	130.50	Matches:	81
Percent Similarity:	34.37%	Conservative:	41
Best Local Similarity:	22.82%	Mismatches:	128
Query Match:	7.81%	Indels:	106
DB:	4	Gaps:	17

US-10-018-706-2 (1-322) x US-09-103-840A-2 (1-4403765)

Qy 40 AenSerThrSerGlySerGlySerHisArgThrSerGlySerGlyGlyLeuAlaIleGly 59

Db 431307 AACTCGGGACCGGACCTGCGGATCGGCAACTCGGACGCGGACACCGCGCATTCGCG 431248

Qy 60 SerGlnValIleThrAsp-----SerGlnValProAsnArgTyrGlnVal 75

Db 431247 AACACCGGACGACCTAACACCGGCTTCTTCAACACCGGACCTGTAACACCGGCTGCGC 431188

Qy 76 LysGlnGlyAspThrValSerLysIleAlaGlnArgTyrGlyLeuAsnTyrArgGluIle 95

Db 431187 AACCGGGGACG-----TACAACACCGGCTGGTGAACACAC 431152

Qy 96 GlyHisIleAsn----- 99

Db 431151 GCGACACCAACACCGGACCTGCGGACCTTCAACACCGGCTTCAACACAC 431092

Qy 100 ---AenLeuAsnSerSerTyrThr-----IleTyrThrGlyGlnTyrLeuThr 114

Db 431091 GGCATTTTCAGTACGCGGCTTCCCAACAGGCTGATATCGCCACCGGCGCTTATCACC 431032

Qy 115 -----LeuTyrSerGlyAspLeuLysValArgGluArgSerIle 127

Db 431031 GCGCATGCGGACACCGGCTTCTGCGGCGGAC-----CACACGCGCTTATTC 430981

Qy 128 SerSerGlyValAsnThr-----AlaHisThrProSerProValAla 141

Db 430980 AGCGGGGCTATCGGGCTCATGTTCCGAATATCCGACACGTCACCGGTGAATTC 430921

Qy 142 ValGln-SerSerArgProProValGlnGlnHisProAlaValGlnLysProThrProPr 161

Db 430920 GTCAACATCCCATACCGGACGCT-----TACACACACCGCTC 430882

Qy 161 ovalValVal-----ValLysLysProThrProThrProProValValGlnProAl 179

Db 430881 TACAGCGCATACCGCTTGAGCAATCAACTCGGCTTTCACCATGACATCGACGAGGATC 430822

Qy 179 aproValAlaProProValThrGlnAlaProPheAlaThrGlySerSerGlyValMetGly 199

Db 430821 CCCCTGCTGGCTGGTGCATATGACGAGCCGCTCTCCGCCCATATCCGGAACCGGTCC 430762

Qy 199 nPheArgTyrProValGlyAlaThr-----AsnProValAlaArgArgPheGlyTh 216

Db 430761 GCGATACCGGTCAACATCGGAGCACCTCGGCGGTTCGACCGCG-ATCAGGATCCGGGCAC 430703

Qy 216 rAlaThrValAlaIleSerThrValThr----- 225

Db 430702 CGCAAGGTCGCTCCCTTCGATGTCACGTTGTCACATTCGCGCTACACGCGCTTTT 430643

Qy 226 -----SerAsnGlyMetTyrPheSerGlyArgAspGlyAspLeuIle 239

Db 430642 CAAGCCACACCGATCCGTCCTCGGCG-----TTCTTCAACGCGCGCCCCGGAACCGTATC 430586

Qy 239 eAsnAlaSerAsnAlaGlyThrValIleGlnAlaAspHisAsnMetAspGlyAlaSerIle 259

Db 430585 GGGCATCGCCAAATCGCGCCCAATTCGCGCTTCGACAGAACGTCGGAATCGCGG-- 430528

Qy 259 eValIleGlnHisThrAsnGlyPheValSerSerTyrIleHisIleLysAspAlaGlnVa 279

Db 430527 -----ACCTCGGCTTC--AACCACTACGCGCTCGCTGCATCGGACCTGCG 430484

Qy 279 lLysThrGlyAspThrVal-----ArgThrGly-----GlnArgGly 291

Db 430483 GAACCTGGCGCATACCGCTTCGCGGCTATTCACACCGGACCTCGGGGACACCGGCAACGT 430424

Qy 291 eAlaSerMetLysAsn-----GlnProSerGly 300

Db 430423 CTCGGGATGTTCAACATCGGACGACCACTCGGCGGTTCTTCCACGACGACGAGGACCG 430364

Qy 300 yAlaAlaLeuPheGluPheArgGlyIleSerArgAsnGlyValTyr 314

Db 430363 GATGTCATGTTCAACTCGGCTCGGGAACATCGGCAATTC 430321

RESULT 9

US-09-103-840A-1/C

Sequence 1, Application US/09103840A

Patent No. 6294328

GENERAL INFORMATION:

APPLICANT: FLEISCHMAN, Robert D.

APPLICANT: WHITE, Owen R.

APPLICANT: FRASER, Claire M.

APPLICANT: VENTER, John C.

TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM

TITLE OF INVENTION: TUBERCULOSIS

FILE REFERENCE: 24366-20007.00

CURRENT APPLICATION NUMBER: US/09/103,840A

CURRENT FILING DATE: 1998-06-24

NUMBER OF SEQ ID NOS: 2

SOFTWARE: Patent In Ver. 2.1

SEQ ID NO 1

LENGTH: 4411529

TYPE: DNA

ORGANISM: Mycobacterium tuberculosis

OTHER INFORMATION: H37Rv

US-09-103-840A-1

Alignment Scores:

Pred. No.:	656	Length:	4411529
Score:	130.50	Matches:	81
Percent Similarity:	34.37%	Conservative:	41
Best Local Similarity:	22.82%	Mismatches:	128
Query Match:	7.81%	Indels:	106
DB:	4	Gaps:	17

US-10-018-706-2 (1-322) x US-09-103-840A-1 (1-4411529)

Qy 40 AenSerThrSerGlySerGlySerHisArgThrSerGlySerGlyGlyLeuAlaIleGly 59

Db 431224 AACTCGGGACCGGACCTGCGGATCGGCAACTCGGACGCGGACACCGGATTCGCG 431165

OY	60	SeGInVal1lleThrsP-----SeGInVal1ProAsnArglyGInVal	75
Db	431164	AACACCGGAGCGACTAACACGGGCTTTTCAACACCGGATCGTCAACACCGGTGTCCG	431105
OY	76	LySGInGlyAspThrValSerIySleIaGlnArglyGlyLeuAsnThrArgGluIle	95
Db	431104	AACGGGGGAGC-----TCAACACCGGCTGTGTCAACAC	431069
OY	96	GlyHISleAsn-----	99
Db	431068	GGGACACCAACACCGGATCGGCAACCTGGGAGCTTCAACACGGGCTTCAACAACC	431009
OY	100	---AspLeuAsnSerIyThr-----IleTyThrGlyInTPleuThr	114
Db	431008	GGCAATTTCAGTACGGGGCTTGTCCAAACAGGTGATATGCCACCGGGGTTTTTCAATCC	430949
OY	115	-----LeuTPSerGlyAspLeuLySleValArgGlnArgSerIle	127
Db	430948	GGCGACATGGCAACCGCGCTTCTGGCGCGGAC-----CAGACGGCTTATTC	430898
OY	128	SerSerGlyValAsnThr-----AlaHisThrProSerProValAla	141
Db	430897	AACGGCGGTATCGGTCCATGTTCCGAAATCCCGCACACGTCAACCTGTGAAGTTCCC	430838
OY	142	ValGln-SerSerArgProProValGlnGlnHisProAlaValGlnLyProThrPro	161
Db	430837	GTCAACATCCCATCAACCCCGACCT-----TCAACCAACCGTC	430799
OY	161	oValValVal-----ValIySerProThrProThrProProValValGlnGlnProAl	179
Db	430798	TACAGCGGATACACGCTTGAGCAATCACTTCGTTCAACATGCACATCGAGGATC	430739
OY	179	AProValAlaProProValThrGlnAlaProPheAlaThrnglySerSerGlyValMetel	199
Db	430738	CCCTCGTCGGCCGGTGCATACAGACGGCTTTCTCCGCCCATCCGGACCGGTCCC	430679
OY	199	nPheArgTyrrProValGlyValAlaThr-----AsnProValValArgArgPheGlyTh	216
Db	430678	GGCATACGGTCAACATCGGCAACCTGGCGGTTGCACCGC-ATCAGATCCCGGCAC	430620
OY	216	rAlaThrValAlaGlySerThrValThr-----	225
Db	430619	CGCAAGCGTGGTCCCTTGATGTACAGTGTCAACATTCGGGCTTACACACGGGCTTTT	430560
OY	226	-----SerAsnGlyMetrPheSerGlyArgAspGlyAspLeuI	239
Db	430559	CAACGCCACACCGATCCGCTCTCGGGC---ATTCTTCAACGGCGGCCCGGAAACGTATC	430503
OY	239	eAsnAlaSerAsnAlaGlyThrValIleGlnAlaAspHisAsnMetAspGlyValSerI	259
Db	430502	GGGCACTGCCAACATCGGGCCAAACATTCGCGCTTCCAGAACGTGCGCAATCCGGC-	430445
OY	259	eValIleGlnHisThrAsnGlyPheValSerSerTyrrIleHisIleTyAspAlaGlnVa	279
Db	430444	-----ACCTGGGCTTC---AACAACTACGGCTCGGTGCAATCGGAGCTGGC	430401
OY	279	ILyThrGlyAspThrVal-----ArgThrGly-----GlnArgI	291
Db	430400	GAACCTGGGCGATACCGTCTGGGGGTATTCAAACACCGGCATCGGGGACACCGCAACT	430341
OY	291	eAlaSerMetLyAsn-----GlnProSerG	300
Db	430340	CTCGGCGATGTTCACATCGGACAGCAACTCGCGGGCTTTTCCACAGACACGGACACCG	430281
OY	300	yAlaAlaLeuPheGlnPheArgIleSerArgAsnGlyValIyrr	314
Db	430280	GATGTGATGTTCACCTCGGCTGTGGAAACATCGCCCAATTC	430238

```

/ GENERAL INFORMATION:
/ APPLICANT: FLEISCHMAN, Robert D.
/ APPLICANT: WHITE, Owen R.
/ APPLICANT: FRASER, Claire M.
/ APPLICANT: VENTER, John C.
/ TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
/ FILE REFERENCE: 24366-20007.00
/ CURRENT APPLICATION NUMBER: US/09/103,840A
/ CURRENT FILING DATE: 1998-06-24
/ NUMBER OF SEQ ID NOS: 2
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO: 1
/ LENGTH: 4411529
/ TYPE: DNA
/ ORGANISM: Mycobacterium tuberculosis
/ OTHER INFORMATION: H37Rv
US-09-103-840A-1

Alignment Scores:
Pred. No.: 3.1e+03 Length: 4411529
Score: 123.50 Matches: 53
Percent Similarity: 34.21% Conservative: 12
Best Local Similarity: 27.89% Mismatches: 74
Query Match: 7.40% Indels: 52
DB: 4 Gaps: 10

US-10-018-706-2 (1-322) x US-09-103-840A-1 (1-4411529)
QY 135 HistHrProSerProValAlaValAlaGlnSerSerThrProProValAlaGlnGlnHisProAla 154
Db 1990636 CACCGCGCGAGCCCA---GCCGCCCGCATAGCAAGCGGGCTTCCCGCGCGCCCG--- 1990689
QY 155 ValGlnHisProThrProProProValValValValHisHisProThr---ProThrProPro 173
Db 1990690 -----CCGAGCGCCCGCGGTGTGTGTGTGGCGGCGCCCGCGGACACCAACCGTACCGCG 1990740
QY 174 -----ValValGlnGlnProAlaProValAlaProProValThrGlnAlaPro 189
Db 1990741 GACCGCATCAAGAAAGCGCGATCGCGCGCGCACCGCGCCCGCGCGCACCGCGCGTTCCG 1990800
QY 190 PheAlaThrGlySerSerSerGlyValMetGlnPheArgTyr-----ProValGly 205
Db 1990801 AC-GCGCGCGGCGCGCGCGCGCGCACCGATGCCAAAGAGATCCCGCGCTGCCACCGGC 1990859
QY 206 Ala-----ThrAsnProValValAlaArgPheGlyThr 216
Db 1990860 GCCCGCGCGCGCTGCGGTGTGTGTGCGCGCGCACCGCGCCCGCGCGTGTGCGCGTGGCGAA 1990919
QY 217 AlaThrValAlaGlySerThrValThrSerAsnGlyMetTrpPheSerGlyArgAspGly 236
Db 1990920 CACCCAGCGCGCGCGCACCGCATC----- 1990943
QY 237 AspLeuIleAsnAlaSerAsnAlaGlyThrValIleGlnAlaAspHisAsnMetAspGly 256
Db 1990944 -----GTCGCCGGTTCGCGCGCGCGCCACCTGTGCGCGTACCGATCAGCGGCGGCTCCGT 1990997
QY 257 AlaSerIleValIleGlnHisThrAsnGlyPheValSerSer-----TyrIle 272
Db 1990998 -----CAATGCGTCGCGGGGTTCGTGATGAACCTGAATGTCGTCGT 1991042
QY 273 HisIleLysAspAlaGlnValGlyHisThrGlyAspThrValAlaArgThrGlyGlnAlaGlieAla 292
Db 1991043 CAGGTGTGCGCATGCGCAGGTGCTCTCGGAGCGCTTATATCCGTCGCGC----- 1991090
QY 293 SerMetLysAsnGlnProSerGlyAlaAla 302
Db 1991091 ---GCCAGCAGCAACCCGCGAAGCGCC 1991117

RESULT 11
US-08-276-213-6
/ Sequence 6, Application US/08276213
/ Patent No. 5536655

```

GENERAL INFORMATION:
APPLICANT: Thomas, Steven
APPLICANT: Laymon, Robert
APPLICANT: Himmell, Michael
TITLE OF INVENTION: GENE ENCODING FOR THE E1 ENDOGLUCANASE
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESS: National Renewable Energy Laboratory
STREET: 1617 Cole Boulevard
CITY: Golden
STATE: CO
COUNTRY: USA
ZIP: 80401-3393
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/276,213
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: O'Connor, Edna
REGISTRATION NUMBER: 29,252
REFERENCE/DOCKET NUMBER: NREL IR# 94-08
TELEPHONE: (303)231-1000
TELEFAX: (303)231-1098
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 3004 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
US-08-276-213-6

Alignment Scores:
Pred. No.: 0.0755 Length: 3004
Score: 120.50 Matches: 96
Percent Similarity: 33.98% Conservative: 27
Best Local Similarity: 26.52% Mismatches: 118
Query Match: 7.22% Indels: 122
Gaps: 20

US-10-018-706-2 (1-322) x US-08-276-213-6 (1-3004)

Qy 27 ThrcysileleuAlaGlyCysAlaSerlyProthrtYraSerThrserySeryly 46
Db 1629 ACGTCCCAACCGCTGCTGT-----ACTCCGCGACGACTACGCGACGCG 1676
Qy 47 SerHisArgThrserGlySerylyCysleuAlaIleGlyserGlnValIleThrAspser 66
Db 1677 TCTACCCGCAACGCTGTCA-----GCGATCCGACCTTCC 1712
Qy 67 GlnGlyValProAsnArgTyrGlnValIleGlnGlyAspThrValSer--LysIleAla 85
Db 1713 CCAACAACATGCGCGCATCTGGAACAGAACTGGGATACCTCTTCAATCAGAACTTG 1772
Qy 86 GlnArgTyrGlyLeuAsnTyr--ArgGluIleGlyHisIleAsn----- 99
Db 1773 CACCGTATGCG-----TGGCGCAATTCGATACGACACTGCAATCCAGACGACGAGA 1826
Qy 100 -----AsnLeuAsnSerSerTyr-----ThrIleTyrThrGly 110
Db 1827 CGTGCCTGAAGACGCTGCTCCGATCCTACCTACGCGACCGCGCAATACGCTGGACACT 1886
Qy 111 --GlnTrrPleuThrIleTrrPser----- 117

Db 1887 TCCAGTGG--ACCTTCTGTCTCTGAAACCCCGATCCGCGACAGACAGAAATTCTCA 1943
Qy 118 -----GlyAspPleuLysValArgGluArgS 126
Db 1944 AGGATGACTGGCAGACGCTGACACAGTAAGACGGCTATCTGCGCCGATCAAGTGT 2003
Qy 126 erile-----SerSerylyVala 132
Db 2004 CGATTTTCGATCTGTGCGCGCGCTGTGATCGCTAGACATCAACCGTCCCGTCCGCT 2063
Qy 132 snThrAlaHisThrProSerProValAlaValGlnSerSerArgProProValGlnGln 152
Db 2064 CGCCGTCCTCCGTCGCGCGCGCTGCGG-----AGTCGACCGCGACGCGCTACTC 2114
Qy 152 isProValAlaGlnLysProThrProProValAlaValIleLysLysProThrProThr 172
Db 2115 CGACGCGCAGACGACGCGCGCGCA-----ACGCTGACCCCTACTGCTACGCGCGC 2168
Qy 172 roProValAlaGlnGlnProAlaProValAla--ProProValThrGluAlaProPheAla 191
Db 2169 CCAAGGCAAGCCGACGCGCTGACAGCGAGCTCCGAGCCCGCTGACCGCGAGT 2228
Qy 192 ThrGlySerSer-----GlyValMetGlnPheArgTyrProValGlyAlaThrAsnPro 209
Db 2229 ACCAGGTCAACAGCGATGGGGCAATGCG--TTCAAGGTAAACGCTGGCCGTGACAAAT--- 2284
Qy 210 ValValArgArgPheGlyThrAlaThrValAlaGlySerThrValThrSerAsnGlyMet 229
Db 2285 -----TCCGATCCGTCGCGACCAAGACATGACGCTGAGT----- 2320
Qy 230 TrpPheSerGlyArgAspGlyAspLeuIleAsnAlaSerAsnAlaGlyThrValIleGln 249
Db 2321 TGAACATTCGCGGGAATCAGACGATACCAATTCGTGATGACGCGTC----- 2371
Qy 250 AlaAspHisAsnMetAspGlyAlaSerIle-----Val 260
Db 2372 -----ACGCAAGACGTCAGTCGCTGGAATGATGATTATACAAACGTCG 2422
Qy 261 IleGln-----HisThrAsnGlyPheValSerSerTyrIle----- 272
Db 2423 ATTCAGCTGCTGCAACACACGCTGCGATTCAGAGGACGACTATACCGGAGCAACGCG 2482
Qy 273 -----HisIleLysAspAla--GlnValIleThrGlyAspThrValArgThrGlyGln-- 289
Db 2483 GCACCGACAGTGGCTGCGCAGCAAGTTAACTGCGGGAGCCGACGCGAGGCTCCGGA 2542
Qy 290 -----ArgIleAlaSerMetLysAsnGlnProSerGlyAlaAla 302
Db 2543 CCGTCCGTTCCCGCGCTTCCACTATGAGGCAACCAACATCCGACGCGAAGTGC 2600

RESULT 12
US-09-136-574A-1
Sequence 1, Application US/09136574A
Patent No. 6294366
GENERAL INFORMATION:
APPLICANT: Farrington, Graham K.
Anderson, Paige
Gibbs, Moreland
Bergquist, Peter
Daniels, Roy
Morgan, Hugh W.
Williams, Diane P.
TITLE OF INVENTION: Compositions and Methods for
Treating Cellulose Containing Fabrics Using Truncated
Cellulase Enzyme Compositions
NUMBER OF SEQUENCES: 49
CORRESPONDENCE ADDRESS:
ADDRESS: Howson and Howson
STREET: Spring House Corporate Center, P.O. Box 457
CITY: Spring House
STATE: PA
COUNTRY: USA
ZIP: 19477

[illegible]

```

QY 92 ----TTPARG-----GulieglyHisle 98
Db 2002 TTAATGCTACACCGCTGAGGAGATTGCTGGGACATCTTCCCGCTTCTCCCGCTGGGACATT 2061
QY 99 -----AaaenLeuAaenSerSerTyThrIleTyThrIleTyThrIleTyThrIle 114
Db 2062 TGGAGAGCGTGTACACCTTTCTGCGGGAGGAGGACCTTGTATACCGGACCTGGTGCAACA 2121
QY 115 LeuTrpSerGlyAsePleuLyValArgGluArgSerIleSerSerGlyValAseThrAla 134
Db 2122 -----TCTGGCTTTTCTTACGGCAT 2139
QY 135 HisThrProSerProValAlaValGlnSerSerArgProProValGlnGlnHisProAla 154
Db 2140 TTCTCCCGCCCT-----GAGGCGGCGCGCCCGCTTTGGCGCGCTGCGCCGCGG 2187
QY 155 ValGlnIlyProThrProProVal----- 162
Db 2188 CTGCGCCACCGCTTACCCACCGTATAGACATTGGGNGTTACACACCACTTAAAGAG 2247
QY 162 ----- 162
Db 2248 TCTCAGGTGATGCGGAGCATGTGTCGCCCTGCTCTGAGCCCGCTGATTAACCACTCC 2307
QY 163 ValValValIlySlyProThrProThrProProValValGlnGlnProAlaProValAla 182
Db 2308 ATATGGCTTACCTCCCTCCCGCCCGCTCTCTCTGTGGGTAGCCACCAACACC----- 2361
QY 183 ProProValThrGluAlaProPheAlaThrGlySerSerGlyValMetGlnPheArgTyR 202
Db 2362 CCGGCT-----TCCGCACTGCTGTCCTCTTACACCTAT 2397
QY 203 ProValGlyAlaThrAenProValValArgArgPheGlyThrAlaThrValAlaGlySer 222
Db 2398 CCCACCGCGCA-----AAGGTGATGCGGGCA 2427
QY 223 ThrValThrSerAenGlyMetTyrPheSerGlyArgAsePlyAsePleuIleAseAlaSer 242
Db 2428 TTGTTTGAATCAGACGTACTG-----CTGGTAAAGCTCA 2466
QY 243 AseAlaGlyThrValIleGlnAlaAsePHisAseMetAsePlyAlaSerIleValIleGln 262
Db 2467 AACCGGCGC-----CACCGCCTGAGAGGTGC-----CTCTGC 2499
QY 263 HisThrAenGlyPheAseSerSerTyThrIleHisIleLyAsePlyAlaGlnValLyThrGly 282
Db 2500 CAT-----GCTTTTACCAACGTTTCCAGAGGGGTTTACCCGACCTGAG 2544
QY 283 AseThrValArgThrGlyGlnArgIleAlaSerMetLyAseGlnPro---SerGlyAla 301
Db 2545 TTCAATTATGCTGAGGGGCTTTCGACGCAATTAACCTTGACCCGCGCCTATATTCAATGCA 2604
QY 302 AlaLeuPheGlnPheArgIleSerArgAse 311
Db 2605 GTGCGCCCGCACTATAGGTTGAGCAGAT 2634
RESULT 15
US-09-462-606-3
Sequence 3, Application US/09462606
Patent No. 6432408
GENERAL INFORMATION:
APPLICANT: MENG, XIANG-JIN
APPLICANT: Emerson, Suzanne U.
APPLICANT: Putcell, Robert H.
TITLE OF INVENTION: A SWINE HEPATITIS B VIRUS AND USES THEREOF
FILE REFERENCE: 202642670US1
CURRENT APPLICATION NUMBER: US/09/462,606
CURRENT FILING DATE: 2000-06-12
PRIOR APPLICATION NUMBER: US 60/053069
PRIOR FILING DATE: 1997-07-18
PRIOR APPLICATION NUMBER: PCT/US98/14665
PRIOR FILING DATE: 1998-07-17

```

NUMBER OF SEQ ID NOS: 65
 SOFTWARE: PatentIn Ver. 2.1
 SEQ ID NO: 3
 LENGTH: 7207
 TYPE: DNA
 ORGANISM: Hepatitis E virus
 US-09-462-606-3

Alignment Scores:
 Pred. No.: 0.511 Length: 7207
 Score: 118.00 Matches: 79
 Percent Similarity: 31.54% Conservative: 44
 Best Local Similarity: 20.26% Mismatches: 129
 Query Match: 7.07% Indels: 138
 Gaps: 17

US-10-018-706-2 (1-322) x US-09-462-606-3 (1-7207)

```

QY 1 MetThrValThrIleAlaIleAsnSerGlnAsnGlnIleProIleLysArgLeuGlyLeu 20
DB 1651 CTAAACGGCTACTGTTGAGCTCACTGCAGAGCCGAGCCGTTTAGAGTCCGCACTGCTT 1710
QY 21 IlePheGlyValIleThrThrCysIleLeuAlaGlyCys-----AlaSerLysPro 37
DB 1711 GGTAATPAGACCTTCAGACGACGCGGTGATGGCGCCCATCTTGAGGCCGAAATGCTCT 1770
QY 38 ThrTYrAsnSerThrSerGlySerHisArgThrSerGlySerGlyValLeuAla 57
DB 1771 GAGCAGTATGTCCTATTCATTCACGCGCTCCGCGCACTATGAGGCGCGGTCACTAC 1830
QY 58 Ile-----GlySerGlnValIleThrAspSerGlnGlyValPro 70
DB 1831 CTCACCTTAGACTACCCCTCGCGCGCTGCGAGGTAGGATTCATCATATGACCTTGAT 1890
QY 71 AsnArgTYrGlnValIleGlnGlyAspThrValSer----- 82
DB 1891 TGCACAGCCACATTCCTCCCGCGCGCGCTTACGGCTGCGCGCGGAGGTGCGCGCC 1950
QY 83 -----LysIleAlaGlnArgTYrGlyLeuAsn----- 91
DB 1951 TTTTGAGTGCCTTTATATAGATATATAGTTCACCGCGGCACTTCGTCGCGGATG 2010
QY 92 ---TPArg-----GluIleGlyHisIle 98
DB 2011 TTATGGCTACACCTGAGGAGATTCGCGCATCTTCCCTTTCCTCCCTGGGACAT 2070
QY 99 -----AsnAsnLeuAsnSerSerTYrThrIleTYrThrGlyGlnIlePheThr 114
DB 2071 TGGAGACCTGTAACCTTCCTGCGGAGGAGGAGGACCTTGTATACCGGACTTGTCACA 2130
QY 115 LeuTPSerGlyAspLeuLysValArgIuArgSerIleSerSerGlyValAsnThrAla 134
DB 2131 -----TCGGGCTTTCTTACGCAT 2148
QY 135 HisThrProSerProValAlaValGlnSerSerArgProProValGlnGlnHisProAla 154
DB 2149 TTTCTCCCCCT-----GAGCGCGCGCGCGCGCTTTGGCGCGCTCCCGCGGG 2196
QY 155 ValGlnLysProThrProVal----- 162
DB 2197 CTGCCCCACCTTACCCGACCTGTATGATGACATTTGGGTTCACACACCTTCAGAGAG 2256
QY 162 ----- 162
DB 2257 TCTCAGTGCATGCGGACCTGTGCCCCCTGCTCTGAGCCCGCTGATACCGAGCTCC 2316
QY 163 ValValValLysLysProThrProThrProProValValGlnGlnProAlaProValAla 182
DB 2317 ATTGGCTTACCTCCCTCCCTCCCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2370
QY 183 ProProValThrGluAlaProPheAlaThrGlySerSerGlyValMetGlnPheArgTYr 202
DB 2371 CCGGCT-----TCCCGCACTCGTGTCTCTCTCTACACCTAT 2406

```

```

QY 203 ProValGlyAlaThrAsnProValValArgArgPheGlyThrAlaThrValAlaGlySer 222
DB 2407 CCCGACGGCGCA-----AAGGTGTATGCGGGGTCA 2436
QY 223 ThrValThrSerAsnGlyMetTrpPheSerGlyArgAspGlyAspLeuIleAsnAlaSer 242
DB 2437 TTGTTGAATCAGACTGTAACCTG-----CTGGTTAATGACCTCA 2475
QY 243 AsnAlaGlyThrValIleGlnAlaAspHisAsnMetAspGlyAlaSerIleValIleGln 262
DB 2476 AACCGCGCG-----CACCGCGCTGAGGCTGCG-----CTCTGC 2508
QY 263 HisThrAsnGlyPheValSerSerTYrIleHisIleLysAspAlaGlnValIleThrGly 282
DB 2509 CAT-----GCTTTTACCAACGTTTCCGAGAGGCGTTTACCGAGCTGAG 2553
QY 283 AspThrValArgThrGlyGlnArgIleAlaSerMetLysAsnGlnPro-----SerGlyAla 301
DB 2554 TTCAATTAGCCTGAGGCGCTTCGACAGCATATACCTGACCGCGCGCTTATTCATTCATCA 2613
QY 302 AlaLeuPheGluPheArgIleSerArgAsn 311
DB 2614 GTGGCCCCGCACTATGAGGTGAGCAGCAAT 2643

RESULT 16
US-09-010-928B-1/C
Sequence 1, Application US/09010928B
Patent No. 5994099
GENERAL INFORMATION:
APPLICANT: Lewis, Randolph V
APPLICANT: Hayashi, Cheryl Y
TITLE OF INVENTION: EXTREMELY ELASTIC SPIDER SILK PROTEIN AND DNA
TITLE OF INVENTION: CODING THEREFOR
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESSES:
ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH
ADDRESS: 8110 GATEHOUSE RD. SUITE 500B
CITY: FALLS CHURCH
STATE: VIRGINIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 22042
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/010,928B
FILING DATE: 22-JAN-1998
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Murphy Jr, Gerald M
REGISTRATION NUMBER: 28997
REFERENCE/DOCKET NUMBER: 1447-109P
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2830 base pairs
TYPE: nucleic acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
FEATURES:
NAME/KEY:
LOCATION: 1..2830 /note= "Flagelliform DNA sequence
OTHER INFORMATION: taken from the 5' region. The putative start codon is at
OTHER INFORMATION: position 219"
NAME/KEY: CDS
LOCATION: 219..2830
US-09-010-928B-1

```

Alignment Scores:

Pred. No.:	0.134	Length:	2830
Score:	117.50	Matches:	49
Percent Similarity:	36.04%	Conservative:	22
Best Local Similarity:	24.87%	Mismatches:	86
Query Match:	7.04%	Indels:	40
DB:	2	Gaps:	7

US-10-018-706-2 (1-3322) x US-09-010-928B-1 (1-2830)

QY 26 ThrThrCysIleLeuAlaGlyCysAlaSerIleProThrTyraSerThrSerGlySer 45
 DB 2219 ACCGACCTGGTCCGAGGTCCACACGACCTGACCGTAGGGTCCACGACGACGAC 2160
 QY 46 GlySerHisArgThrSerGlySerGlyGlyLeu-----AlaIleGlySerGlnValIle 63
 DB 2159 GGGTCTCCGAAACCGGGGTCTCCGAAACGACGACGACGACGACGACGACGACG 2100
 QY 64 ThrAspSerGlnGlyValProAsnArgTyrglnValIleGlnGlyAspThrValSerIys 83
 DB 2099 TCCGCACTTCGGGTCCGAGACGACGACGACGACGACGACGACGACGACGACG 2040
 QY 84 IleAlaGlnArgTyrglyLeuAsnTrp-----ArgGlnIleGlyHisIleAsn 99
 DB 2039 TACGCTCCAGTCTCTACGCGACGACGACGACGACGACGACGACGACGACG 1980
 QY 100 AsnLeuAsnSerIleTyrrThrIleTyrrThrGly----- 110
 DB 1979 GGAGCCACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 1920
 QY 111 -----GlnTrp-----LeuThrLeuTrpSer 117
 DB 1919 AAAGCTACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 1860
 QY 118 GlyAspLeuIysValArgGlnArgSerIleSerGlyValAlaThrAlaHisThr-Pr 137
 DB 1859 -----AGAGCGACGACGACGACGACGACGACGACGACGACGACGACGACG 1821
 QY 137 oSerProValAlaValGlnSerSerArgProProValGlnGlnHisProAlaValGlnIly 157
 DB 1820 GGGTCCGAAACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 1761
 QY 157 sProThrProProValValValValIlyLysProThrPro-----ThrProProValIva 175
 DB 1760 GCCATCTGCACATCAATGTATATATCCAAATCTCTACGATGTGTGCGCGCG- 1706
 QY 175 lGlnGlnProAlaProValAlaProProValThrGlnAlaProPhaIa 191
 DB 1705 -GATCTCTGCTGCTCACCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1658

RESULT 17

US-09-125-287-2

Sequence 2, Application US/09125287B

Patent No. 6114602

GENERAL INFORMATION:

APPLICANT: BARG, Rivka

APPLICANT: SALTS, Tehiam

TITLE OF INVENTION: METHOD FOR THE INTRODUCTION OF GENETIC PARTHENOCAPI IN

TITLE OF INVENTION: PLANTS

FILE REFERENCE: INTRO GENETIC PARTHENOCAPI IN PLANTS

CURRENT APPLICATION NUMBER: US/09/125,287B

CURRENT FILING DATE: 1998-11-09

EARLIER APPLICATION NUMBER: PCT/IL97/00051

EARLIER FILING DATE: 1997-02-13

NUMBER OF SEQ ID NOS: 6

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 2

LENGTH: 4518

TYPE: DNA

ORGANISM: TPRP-F1 PROMOTOR

FEATURE:

NAME/KEY: unsure

LOCATION: (6)..(191)

OTHER INFORMATION: "n"s are any nucleic residue

US-09-125-287-2

Alignment Scores:

Pred. No.:	0.387	Length:	4518
Score:	116.00 <td>Matches:</td> <td>35</td>	Matches:	35
Percent Similarity:	44.76% <td>Conservative:</td> <td>12</td>	Conservative:	12
Best Local Similarity:	33.33% <td>Mismatches:</td> <td>40</td>	Mismatches:	40
Query Match:	6.95% <td>Indels:</td> <td>19</td>	Indels:	19
DB:	3	Gaps:	5

US-10-018-706-2 (1-3322) x US-09-125-287-2 (1-4518)

QY 136 ThrProSerProValAlaValGlnSerSerArgProProValGlnGlnHisProAlaVal 155
 DB 3313 ACTCAACACCACTTATGTTTCCACCCGACGACGACGACGACGACGACGACGACG 3357
 QY 156 GlnIysProThrProProValAla-----ValAlaIlyLysPro-----ThrProThr 171
 DB 3358 CCATCAACCAACCTCTATGTTTCCACCCGACGACGACGACGACGACGACGACGACG 3417
 QY 172 ProProValValGlnGlnProAlaProAlaProProValThrGlnAlaProPhaIa 191
 DB 3418 CCACCTGTGTGTCACCTCAATCATTCACCAACGACGACGACGACGACGACGACG 3477
 QY 192 ThrGlySerSerGlyValMetGlnPheArgTyrr----- 202
 DB 3478 CCATCTCTCCGCTGTAATACCAACGACGACGACGACGACGACGACGACGACG 3537
 QY 203 ProValGlyAlaThrAsnProProValAlaArgPheGlyThrAlaThrValAlaGlySer 222
 DB 3538 CCATCACT-TCCAACACCCCTTACCATGCCACGACGACGACGACGACGACGACG 3596
 QY 223 ThrValThrSerAsn 227
 DB 3597 ACCATCAACCAACGACG 3611

RESULT 18

US-09-125-287-1

Sequence 1, Application US/09125287B

Patent No. 6114602

GENERAL INFORMATION:

APPLICANT: BARG, Rivka

APPLICANT: SALTS, Tehiam

TITLE OF INVENTION: METHOD FOR THE INTRODUCTION OF GENETIC PARTHENOCAPI IN

TITLE OF INVENTION: PLANTS

FILE REFERENCE: INTRO GENETIC PARTHENOCAPI IN PLANTS

CURRENT APPLICATION NUMBER: US/09/125,287B

CURRENT FILING DATE: 1998-11-09

EARLIER APPLICATION NUMBER: PCT/IL97/00051

EARLIER FILING DATE: 1997-02-13

NUMBER OF SEQ ID NOS: 6

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 1

LENGTH: 12839

TYPE: DNA

ORGANISM: TPRP-F1 GENOMIC CLONE

FEATURE:

NAME/KEY: unsure

LOCATION: (5)..(11322)

OTHER INFORMATION: "n"s are any nucleic residue

US-09-125-287-1

Alignment Scores:

Pred. No.:	1.95	Length:	12839
Score:	116.00 <td>Matches:</td> <td>35</td>	Matches:	35
Percent Similarity:	44.76% <td>Conservative:</td> <td>12</td>	Conservative:	12
Best Local Similarity:	33.33% <td>Mismatches:</td> <td>40</td>	Mismatches:	40
Query Match:	6.95% <td>Indels:</td> <td>19</td>	Indels:	19
DB:	3	Gaps:	5

US-10-018-706-2 (1-3322) x US-09-125-287-1 (1-12839)

QY 136 ThrProSerProValAlaValAlaGlnSerSerArgProProValAlaGlnHisProAlaVal 155
 DB 3314 ACTCCAAACACCACTATTATGTC-----CATCCACCACTACTCCAAACCA----- 3358
 QY 156 GlnLysProThrProProValAla-----ValValLysLysPro---ThrProThr 171
 DB 3359 CCATCAACCAACACCTCATTTGTTTCACCCCCCATTTGTTTATCCACCATCACTCCACCA 3418
 QY 172 ProProValAlaGlnGlnProAlaProValAlaProProValAlaGlnLysProPheAla 191
 DB 3419 CCACCTGTTGTCTACCTCCAAATCAATCCAAACACCACTATGTCCTCCACCTTTGTC 3478
 QY 192 ThrGlySerSerGlyValMetGlnPheArgTyr----- 202
 DB 3479 CCCAATCTCTCCGTGTATATACCACCACTCACTGTCGCAAGTCTCCGGTGTACTCCA 3538
 QY 203 ProValAlaValAlaThrAsnProValAlaValArgArgPheGlyThrAlaThrValAlaGlySer 222
 DB 3539 CCATAGT-TCCAAACACCCCTTACACCATGCCACACCAACACCAACCAACATATAT 3597
 QY 223 ThrValThrSerAsn 227
 DB 3598 ACCATCAACCAACG 3612

RESULT 19

US-08-194-290-6
 ; Sequence 6, Application US/08194290

Patent No. 5500353
 ; GENERAL INFORMATION:

APPLICANT: Smtc, John

APPLICANT: Bingle, Wade H

TITLE OF INVENTION: Bacterial surface protein expression

NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:

ADDRESSEE: Shlesinger, Arkwright & Garvey

CITY: Arlington

STATE: Virginia

COUNTRY: USA

ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/194,290

FILING DATE: 09-FEB-1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Garvey, George A

REGISTRATION NUMBER: 17737

REFERENCE/DOCKET NUMBER: 5946-1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 703-684-5600

TELEFAX: 703-836-5288

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 3300 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

ORIGINAL SOURCE:

ORGANISM: Caulobacter crescentus

STRAIN: CB 15

US-08-194-290-6

Alignment Scores:

Pred. No.: 0.297 Length: 3300
 Score: 115.00 Matches: 75
 Percent Similarity: 36.48% Conservative: 41

Best Local Similarity: 23.58% Mismatches: 134
 Query Match: 6.89% Indels: 69
 DB: 1 Gaps: 12

US-10-018-706-2 (1-322) x US-08-194-290-6 (1-3300)

QY 23 GlnValIleThrThrCysIleLeuAlaGlyCysAlaSerLysProThrTyrAsnSerThr 42
 DB 776 GGGCGCCGTGCGACCGACCAACCGCGCTGGCTGGAACCTGTTCACCGCTTACCGCTGCG 835
 QY 43 SerGlySerGlySerHisArgThrSerGlySerGlyLeuAlaIleGlySerGlnVal 62
 DB 836 GCGGTGCGGGTTCG-----ACCTTCGCTGACACCGCGACCGACAC 880
 QY 63 IleThrAspSerGlnGly-----ValProAsnArgTyrGln 74
 DB 881 CTGACGGGCGACCGCAACCAACGACGCTGCTGGGTGAAGTCGCGCGCTGCGAC 940
 QY 75 ValLysGlnLysAspThrValSerLysIleAlaGlnArgTyrGlyLeuAsnTrpArgGlu 94
 DB 941 CTGACCGTGGCGACACCTTACGCGCGCTGCTGACCGACGCTGTAACCTGGGTGCA 1000
 QY 95 IleGlyHisIleAsnAsnLeuAsnSerSerTyrThrIleTyrThrGlnGlnTrpLeuThr 114
 DB 1001 GCTGCTGCGGTTCAGGCTCTCCGACCGCGCTGACGATC----- 1039
 QY 115 LeuTrpSerGlyAspLeuLysValArgGlnArgSerIleSerSerGlyValAsnThrAla 134
 DB 1040 -----TCGGGCAATGAACGATG 1057
 QY 135 HisThrProSerProValAlaValAlaGlnSer-SerArg-----Pr 147
 DB 1058 AACGTGACGTGGGCGCTGCGCATCACTGAACACGCTTCGGCGTGAACGCTGAC 1117
 QY 147 OProValGlnGlnHisProAlaValAlaGln---LysProThrProProValAlaValLys 166
 DB 1118 GCCCTGAACACCAACACGCGCGCGCTCAACACCTGACCGCGCGCTGCGCGACG 1177
 QY 166 sLysProThrProThrProProValAlaGlnGlnProAlaProAlaProAlaProAla 186
 DB 1178 CTGACCG---CCACGACCGCGCTC-----AACCGCGAACAACGTCGCGTGA 1228
 QY 186 GlnLysAlaProPheAlaThrGlySerSerGlyValMetGlnPheArgTyrProValAla 206
 DB 1229 -CGGCGCAACGTCAACCGCGCTGCGACGCGGTGACCTCGGCGACGACGACGCG 1287
 QY 206 aThrAsnProValAlaValArgArgPheGlyThrAlaThrValAlaGlySerThrVal 226
 DB 1288 CACTCGCGCGCTTCGCGCACCGGTGCGTGAAGGTGCGGACGACGACGACGACGAC 1347
 QY 226 rAsnGlyMetTrpPheSerArgLysArgAspGlyAspLeuIle-----AspAl 241
 DB 1348 GGGCGCTATCGCGCTGACCGGTGACGCGGTGACCGGTGACCGGTGACCGGTGAC 1407
 QY 241 aSerAsnAlaGlyThrValIleGlnAlaAspHisAsnMetAspGlyAlaSer----- 258
 DB 1408 CGTGAACAC---ACGTGACGCAAGCGGACGTAACCGGTGACCGGTGACGAC 1464
 QY 259 ---IleValIleGlnHisThr-----AsnGlyPheVal 268
 DB 1465 GCGCGTGAACGTCACCAACCGCGCGCGCGCGCGCGCGGTACGCTGCGGTGCG 1524
 QY 268 lSerSerTyrIleHisIleLysAspAlaGlnValLysThrGlyAspThrValArgThr 288
 DB 1525 CAACGCGCTGTGACGATCAACCGGACTCT-----GCCGCGCGCTGCGCGACGCG 1575
 QY 288 YGlnArgIleAlaSerMetLysAsnGlnProSerGlyValAlaLeuPheGlu 305
 DB 1576 CGGCAAGATGCGACGCTGCGGAGGCTTGGCGCGCGCGCGACGATGAC 1627

RESULT 20

US-08-614-377A-6
 ; Sequence 6, Application US/06614377A

Patent No. 5976864
GENERAL INFORMATION:
APPLICANT: Smit, John
APPLICANT: Bingle, Wade H.
TITLE OF INVENTION: EXPRESSION AND SECRETION OF
TITLE OF INVENTION: HETEROLOGOUS
POLYPEPTIDES FROM CAULOBACTER
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson PC
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/614,377A
FILING DATE: 12-MAR-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/194,290
FILING DATE: 09-FEB-1994
CLASSIFICATION: 435
APPLICATION NUMBER: US 07/895,367
FILING DATE: 09-JUNE-1992
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Tsao, Y. Rocky
REGISTRATION NUMBER: 34053
REFERENCE/DOCKET NUMBER: 08106/002001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-542-5070
TELEFAX: 617-542-8906
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 3300 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Caulobacter crescentus
STRAIN: CB 15
US-08-614-377A-6
Alignment Scores:
Pred. No.: 0.297 Length: 3300
Score: 115.00 Matches: 75
Percent Similarity: 36.48% Conservative: 41
Best Local Similarity: 23.58% Mismatches: 134
Query Match: 6.89% Indels: 69
DB: 2 Gaps: 12
US-10-018-706-2 (1-322) x US-08-614-377A-6 (1-3300)
QY 23 GlyValIleThrTrpCysIleLeuAlaGlyCysAlaSerLysProThrTyrAsnSerThr 42
DB 776 GGGCGCCCTGTGACGACGACGCGGCTGCGTGAACCTGTTCCACCGCTATCCGTGTCG 835
QY 43 SerGlySerGlySerHisArgThrSerGlySerGlyGlyLeuAlaIleGlySerGlnVal 62
DB 836 GGGCGGTGGGGTGC-----ACCTCTGCTGACCAACCGGACCGACACC 880
QY 63 IleThrAspSerGlnGly-----ValProAsnArgTyrGln 74
DB 881 CTGACGGGACCGCCACACGACGACGTTCTGTCGGGTGAAGTGCAGCGGCGGACGACC 940

QY 75 ValIysGlnGlyAspThrValSerLysIleAlaGlnArgTyrGlyLeuAsnTrpArgLys 94
DB 941 CTGACCGTGGCCGACACCTGAGGCGGCTGTGACACCGACCTCTGAACCTGGGTGCA 1000
QY 95 IleGlyHisIleAsnAsnLeuAsnSerSerTyrThrIleTyrThrGlyGlnTrpLeuThr 114
DB 1001 GCTGCTCGGCTGACGCTGCTGCGACCGCGGCTGACGATC----- 1039
QY 115 LeuTrpSerGlyAspLeuLysValArgGlnArgSerIleSerSerGlyValAsnThrAla 134
DB 1040 -----TGGGATCGAAGACATG 1057
QY 135 HisThrProSerProValAlaValGlnSer-SerArg-----Pr 147
DB 1058 AACGTGACGTGGCGGCTGGGATCACCTGTAACGCTTCCGGCGTGCAGCGGTGACC 1117
QY 147 OProValGlnGlnHisProAlaValGln--LysProThrProProValValVally 166
DB 1118 GCCCTGAACACCAACACAGCGGCGGCTCAACCGCTCACCGCGCGGTGCAGAAC 1177
QY 166 sLysProThrProThrProProValValGlnGlnProAlaProValAlaProValTh 186
DB 1178 CTGACCG--CCAGACCGCGCTC-----AACCCGGAACAACTGCGCGTGCAGCGG 1228
QY 186 rGluAlaProPheAlaThrGlySerSerGlyValMetGlnPheArgTyrProValGlyAl 206
DB 1229 -CGGCCAAGTCACCGCTGCTGACGGGGGTGTAAGCTCGGCGACGACGCTGCGGCG 1287
QY 206 aThrAsnProValValArgArgPheGlyThrAlaThrValAlaGlySerThrValThrse 226
DB 1288 CAACTCGCGCGCTTCGGGACCGGTGCGTGAAGCGTGCAGAACCTGACGACGACACAC 1347
QY 226 rAsnGlyMetTrpPheSerGlyArgAspGlyAspLeuIle-----AsnAl 241
DB 1348 GGGCGCTATCGCGCTGACCGGTGTGACGCGCGGTGACCGCTGACCGCGCGGACGCG 1407
QY 241 aSerAsnAlaGlyThrValIleGlnAlaAspHisAspMetAspGlyAlaSer----- 258
DB 1408 CGTGAACACC--ACGTGAACGACGACGACGTAACCGGTGACCGGTGACCGACAC 1464
QY 259 ---IleValIleGlnHisThr-----AsnGlyPheVa 268
DB 1465 GGGCGTACGCTGACCCAAACCGCGCGGACCGCGCGCTACCGGTGCGCGGTGCGGT 1524
QY 268 lSerSerTyrIleHisIleLysAspAlaGlnValIleThrGlyAspThrValArgThrG 288
DB 1525 CAACGGCGCTGTGACGATCACCGACTCT-----GCCGCGCTGCGGCGACGACCGC 1575
QY 288 yGlnArgIleAlaSerMetLysAsnGlnProSerGlyAlaAlaLeuPheGlu 305
DB 1576 CGGCAAGATCCGACCGCTGACCGCTGCGGCGCGGCGCGGCGGCGGCGGCGGCGG 1627
RESULT 21
US-09-142-648B-6
Sequence 6, Application US/09142648B
Patent No. 6210948
GENERAL INFORMATION:
APPLICANT: Smit, John
APPLICANT: Bingle, Wade H.
TITLE OF INVENTION: EXPRESSION AND SECRETION OF HETEROLOGOUS POLYPEPTIDES FROM CAULO
FILE REFERENCE: 08106/002002
CURRENT APPLICATION NUMBER: US/09/142,648B
CURRENT FILING DATE: 1999-03-30
PRIOR APPLICATION NUMBER: PCT/CA97/00167
PRIOR FILING DATE: 1997-03-10
PRIOR APPLICATION NUMBER: US 07/614,377
PRIOR FILING DATE: 1996-03-12
SOFTWARE: FastSeq for Windows Version 4.0
NUMBER OF SEQ ID NOS: 12
SEQ ID NO 6
LENGTH: 3300
TYPE: DNA

```

/ ORGANISM: Caulobacter crescentus
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (101)...(3178)
US-09-142-648B-6

Alignment Scores:
Pred. No.: 0.297 Length: 3300
Score: 115.00 Matches: 75
Percent Similarity: 36.48% Conservative: 41
Best Local Similarity: 23.58% Mismatches: 134
Query Match: 6.89% Indels: 69
DB: 4 Gaps: 12

US-10-018-706-2 (1-322) x US-09-142-648B-6 (1-3300)

QY 23 GYValIleThrThrcysIleLeuAlaGlyCysAlaSerLysProThrTyraSerThr 42
DB 776 GGGCGCCCTGTCGACCGAACACGGCGCTGGCGTGAACCTTTCACCCGCTATCCGTCGCG 835
QY 43 SerGlySerGlySerHisArgThrSerGlySerGlyLeuAlaIleGlySerGlnVal 62
DB 836 GGGCGTCCGCTTCG-----ACCTTCGCTGACCAACCGGACCGACAC 880
QY 63 IleThrAspSerGlnGly-----ValProAsnArgTyrGln 74
DB 881 CTGACCGGACCGCCCAACACACACGCTTCGCGGCGTGAAGTCGCGCGCTCGACCC 940
QY 75 ValLysGlnGlyAspThrValSerLysIleAlaArgTyrGlyLeuAsnTyrArgGln 94
DB 941 CTGACCGTGGGACACCTCGACGGCGGTGCTGACACCGACGCTCTGAACCTGGGTGCA 1000
QY 95 IleGlyHisIleAsnAsnLeuAsnSerSerTyrThrIleTyrThrGlyGlnTyrLeuThr 114
DB 1001 GCTGTCGCGTTACGGCTCTGCGACCGCGCGTGCACATC----- 1039
QY 115 LeuTyrSerGlyAspLeuLysValArgIuArgSerIleSerSerGlyValAsnThrAla 134
DB 1040 -----TCGGGCATCGAACAAGAG 1057
QY 135 HisThrProSerProValAlaValGlnSer-SerArg-----Pr 147
DB 1058 AACGAGACCTCGCGCGCTCGATCACCTCGAACACGCTTCGCGCGCTGACGGGTGACCC 1117
QY 147 GProValGlnGlnHisProAlaValGln---LysProThrProProValAlaValVal 166
DB 1118 GCCCTGAACACCAACACACGCGCGGCTCAAAACGTCACCGCGCGGCGCTGACCAAGAC 1177
QY 166 sLysProThrProThrProProValAlaGlnGlnProAlaProValAlaProProValTh 186
DB 1178 CTGACCG---CCACGACCGCGCTC-----AAGCGGCAACACAGCTCCGCTGACAGGG 1228
QY 186 rGluAlaProPheAlaThrGlySerSerGlyValMetGlnPheArgTyrProValGlyAl 206
DB 1229 -CGCGGCAACGTACCGCGCTCGACGGCGCTGACGGCGTGGGACGACACACGCGCGCC 1287
QY 206 rThrAsnProValAlaArgArgPheGlyThrAlaThrValAlaGlySerThrValThse 226
DB 1288 CAACCTCGCGCGCTTCGGGACCGGTGCTGGTGAAGCGTGCAGAACTCGAGACACCAACCC 1347
QY 226 rAsnGlyMetTyrPheSerGlyArgAspGlyAspLeuIle-----AsnAl 241
DB 1348 GGGCGCTATCGCGGTACCGGTGTAACCGCGTGAACCGGTGCTCAAAACGCGCGGCAACCC 1407
QY 241 sSerAsnAlaGlyThrValIleGlnAlaAspHisAsnMetAspGlyAlaSer----- 258
DB 1408 CTGTGAACACC---ACGTTGACCGCAAGCCGACGTGAACCGGTGAACCTCGACGACAC 1464
QY 259 ---IleValIleGlnHisThr-----AsnGlyPheVal 268
DB 1465 GCGCGCGCGGTACCGCAACCGCGCGCGCGCGCGCGCGGTACGCGTGGCGGTGCGGT 1524
QY 268 lSerSerTyrIleHisIleLysAspAlaGlnValLysThrGlyAspThrValArgThrcl 288

```

```

DB 1525 CAACGGCGCTGTGACGATCACCAGCTCT-----GCCGCCGCTGGGCCACGACCGC 1575
QY 288 yGlnArgIleAlaSerMetLysAsnGlnProSerGlyAlaAlaLeuPheGln 305
DB 1576 CGGCAAGATCGCCACGCGTACCTGGGCGAGCTTCGCGCGCCGCCACGATCGAC 1627

RESULT 22
US-09-311-626B-15
/ Sequence 15, Application US/09311626B
/ Patent No. 6399347
/ GENERAL INFORMATION:
/ APPLICANT: Jorgensen, Per Lina
/ APPLICANT: Schmoor, Kirk
/ APPLICANT: Andersen, Lene No. 6399347boe
/ APPLICANT: Schuelein, Martin
/ APPLICANT: Outtrup, Helle
/ TITLE OF INVENTION: No. 6399347el Rhamnogalacturonan Hydrolases
/ FILE REFERENCE: 5572,204-US
/ CURRENT APPLICATION NUMBER: US/09/311,626B
/ PRIOR FILING DATE: 1998-05-13
/ PRIOR APPLICATION NUMBER: 0608/98
/ PRIOR FILING DATE: 1998-05-05
/ NUMBER OF SEQ ID NOS: 61
/ SOFTWARE: FASTSEQ for Windows Version 4.0
/ SEQ ID NO 15
/ LENGTH: 1896
/ TYPE: DNA
/ ORGANISM: Streptomyces coelicolor
US-09-311-626B-15

Alignment Scores:
Pred. No.: 0.176 Length: 1896
Score: 113.50 Matches: 68
Percent Similarity: 35.04% Conservative: 21
Best Local Similarity: 26.77% Mismatches: 99
Query Match: 6.80% Indels: 66
DB: 4 Gaps: 15

US-10-018-706-2 (1-322) x US-09-311-626B-15 (1-1896)

QY 38 ThrTyraSerThrSer-----GlySerGlySerHisArgThrSer 51
DB 1075 ACCAATCTCTCACCACACAGCGGCAAGGCTACGACGGCGGCAACACACAGCTCTCC 1134
QY 52 -----GlySerGlyLeuAlaIleGlySerGlnValIleThr 64
DB 1135 GTCCGCGACGTGACGCGTGAACGCGCGGACGACATCTGACGGCGATGCGCTGAC 1194
QY 65 AspSerGlnGlyValProAsnArgTyrGlnVal-----LysGlnGlyAsp 79
DB 1195 GAC-----AACGGCTACCGCTGTGAGACCAACAGAACGCGGACGCGAC 1239
QY 80 ThrVal-----SerLysIleAlaGlnArgTyrGlyLeuAsnTyrPArgGlnIleGly 96
DB 1240 GCCATGACGTGCGGACCTGACCCGCTCCGGCGGCTGTGAGAGATTCAAGTTCAC 1239
QY 97 HisIleAsnAsnLeuAsnSerSerTyr-----ThrIleTyrThrGlyGln-----Tyr 112
DB 1300 GAGGACGGCTGAAGCCCTCGTACCTGTGCGGACCGCGGACGCGGCGCAATCTCTCG 1359
QY 113 LeuThr-----LeuTyrSerGlyAsp 119
DB 1360 TCCACCGCGCGGACGCGGACACACGCGCGCGTCTCCGGGACATCTGTCGCGAGC 1419
QY 120 LeuLysValArgGln---ArgSerIleSerSerGlyValAsn-ThrAlaHisThrProse 138
DB 1420 GCGGCGCGCGAGTCTGCTGCTCGCGAGAGCGGCATCCGCAACCCCAAGGACCGCTC 1479
QY 138 rProValAlaValGln-----SerSerArgProProValGlnGlnHisProAl 154

```



```

1 PRIOR APPLICATION DATA:
2 APPLICATION NUMBER: JP 224104/1999
3 FILING DATE: 26-ANV-1996
4 PRIOR APPLICATION DATA:
5 APPLICATION NUMBER: JP 48101/1997
6 FILING DATE: 03-MAR-1997
7 ATTORNEY/AGENT INFORMATION:
8 NAME: DAVIDSON, CLIFFORD M.
9 REGISTRATION NUMBER: 32,726
10 REFERENCE/DOCKET NUMBER: 382.1009
11 TELECOMMUNICATION INFORMATION:
12 TELEPHONE: (212)-768-3800
13 TELEFAX: (212)382-2124
14 INFORMATION FOR SEQ ID NO. 9:
15 SEQUENCE CHARACTERISTICS:
16 LENGTH: 2186 base pairs
17 TYPE: nucleic acid
18 STRANDEDNESS: double
19 TOPOLOGY: linear
20 MOLECULE TYPE: DNA (genomic)
21 HYPOTHEetical: NO
22 ANTI-SENSE: NO
23 ORIGINAL SOURCE:
24 ORGANISM: STREPTOMYCES PLATENSIS
25 STRAIN: Q268
26 FEATURE:
27 NAME/KEY: CDS
28 LOCATION: 1477..1911
29 IS-OR-878-546-9

```

LOCATION: replace(1381, "c")
 FEATURE:
 NAME/KEY: variation
 LOCATION: replace(1566, "c")
 FEATURE:
 NAME/KEY: variation
 LOCATION: replace(7010, "g")
 US-07-945-283-1

Alignment Scores:

Pred. No.: 2.22 Length: 8438
 Score: 112.50 Matches: 40
 Percent Similarity: 41.53% Conservative: 9
 Best Local Similarity: 33.90% Mismatches: 58
 Query Match: 6.74% Indels: 12
 Gaps: 3

US-10-018-706-2 (1-322) x US-07-945-283-1 (1-8438)

QY 125 ArgSerIleSerSerGlyValAlaThrAlaHisThr-----ProSerProValAla 141
 Db 6438 CGGGCTTCATCAGGGCTCCGAGCAGGTCACCTCTCTCGGGCCCGGCTCCCGGGCC 6379
 QY 142 ValGlnSerSerArgProProVal-GlnGlnAlaProAlaValGlnLysProThrPro 161
 Db 6378 GGCTCCGGGCGCCGCGCGGCGGACCCCGGCGGACACTCCGAGCGGAGAGCGGGGCGCC 6319
 QY 161 oValValValValValValValValValValValValValValValValValValVal 180
 Db 6318 GGGGGGCGGGGCGGGGCGGGGCGGGGCGGGGCGGGGCGGGGCGGGGCGGGGCGGGGCG 6259
 QY 180 oValAlaProProValThrglnAlaProPheAlaThrglySerSerGlyValMetGln 200
 Db 6258 TCCGGCTCTCCGGCTCCCGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCG 6207
 QY 200 eArgTyProValGlyAlaThraProVal-ValArgArgPheGlyThraAlaThrVala 220
 Db 6206 -----CCCGAGGGGAGGAGCGGCTCGTACGCGGCGGCGGCGGCGGCGGCTCGGAGC 6154
 QY 220 laGlySerThraValThraSerAlaGlyMetThraPheSerGlyValArgGly 236
 Db 6153 CCGGCGGCGGAGCAGCAGCAGCAGCGG-TCGCGGAGCGGCGGCGGCGGCGGCGGCGGCG 6105

RESULT 26

US-08-923-137-2/c
 ; Sequence 2, Application US/08923137
 ; Patent No. 6083716

GENERAL INFORMATION:
 APPLICANT: Wilson, James M.
 APPLICANT: Farina, Steven F.
 APPLICANT: Fisher, Krishna J.
 TITLE OF INVENTION: Chimpanzee Adenovirus Vectors
 NUMBER OF SEQUENCES: 5
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Howson and Howson
 STREET: Spring House Corporate Cntr., P.O. Box 457
 CITY: Spring House
 STATE: Pennsylvania
 COUNTRY: United States of America
 ZIP: 19477

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/923.137

FILING DATE:

CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 60/024,700

FILING DATE: 06-SEP-1996

ATTORNEY/AGENT INFORMATION:

NAME: Bak, Mary E.
 REGISTRATION NUMBER: 31,215
 REFERENCE/DOCKET NUMBER: GNVN.021CPIUSA
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 215-540-9200
 TELEFAX: 215-540-5818

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:
 LENGTH: 36519 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: unknown
 MOLECULE TYPE: cDNA
 US-08-923-137-2

Alignment Scores:

Pred. No.: 23.9 Length: 36519
 Score: 112.00 Matches: 49
 Percent Similarity: 39.79% Conservative: 27
 Best Local Similarity: 25.65% Mismatches: 83
 Query Match: 6.71% Indels: 33
 Gaps: 6

US-10-018-706-2 (1-322) x US-08-923-137-2 (1-36519)

QY 26 ThrThrCysIleLeuAlaGlyCysAlaSerIleProThrThraSer-----ThrSerGly 44
 Db 7308 TCACGATGTACTTCTTGGCTTGGCCACCCCGCTTCAAGAACGACGAGCTTCCAGC 7249
 QY 45 SerGlySerIleArgThrSerGlySerGlyLeuAlaIleGlySerGlnValIleThr 64
 Db 7248 TCTGGAGGAGG-----GGGGCTGGCAGATTCGATTCAGACAGT 7204
 QY 65 AspSerGlnIleValProAlaArgTyArgInValIleGlnIleAspThrValSerIle 84
 Db 7203 TCCTCAGGTCATGTCAGGACGACCTTCCCTCAGGAC-ACCTCCCTGGCGGAGGCC 7145
 QY 85 AlaGlnArgTyGlyLeu-AspThrArgIleIleGlyIleIleAsnAsnLeuAsnSer 104
 Db 7144 GCCAGGCTTACCCCTTCCCGGAGAGGGCTGCGCCCTTCAAGGCGGCTACACAG 7085
 QY 104 rTyThr-----IleTyThrGly-----110
 Db 7084 TTCTACATGCTAGGCTCTTACCGGAGATCAGAGCGGTTCCTCCGAGAGTACTGG 7025
 QY 111 -----GlnTrpLeuThrIleuTrpSerGlyAspLeuValArgIle 125
 Db 7024 AAGACCGGAGAGAGTCTCTTCAACCGGAGCTGTGAAC-----AA 6980
 QY 125 gSerIleSerSerGlyValAsnThrAlaHisThrProSerProValAlaValGlnSer 145
 Db 6979 AAGGACAGCTCAAGATGATCATCATCAGAGACCTTCACACTGCGCTCGAGATC 6920
 QY 145 rArgProProValGlnGlnIleProAlaValGlnIleProThrProProValValVal 165
 Db 6919 CTAGTACACCGGAGCTCGGCGGAGGCTGCAAGACTGCAAGCTTCACTTCACTCGGAGC 6860
 QY 165 llyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllyllylly 185
 Db 6859 TCGGTGGAGCTGGCCAGCGCCACTTCAACATCTTCCAAAGGCGCCACCATCTCTCCA-- 6802
 QY 185 lThrglnAlaProPheAlaThraGlySerSer 195
 Db 6801 -ACTCGATGCCATTTCGCGCAGATCGTCT 6772

RESULT 27

US-09-071-035-221
 ; Sequence 221, Application US/09071035
 ; Patent No. 6448043

GENERAL INFORMATION:
 APPLICANT: Gil H. Choi
 TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
 NUMBER OF SEQUENCES: 496

Wed Jul 9 10:01:00 2003

us-10-018-706-2.rni

Page 20

```

CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
OPERATING SYSTEM: HP Vectra 486/33
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/071,035
FILING DATE:
CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER:
ATTORNEY/AGENT INFORMATION:
NAME: A. Anders Brookes
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB369P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 221:
SEQUENCE CHARACTERISTICS:
LENGTH: 867 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-09-071-035-221

Alignment Scores:
Pred. No.: 0.0912 Length: 867
Score: 111.00 Matches: 75
Percent Similarity: 34.77% Conservative: 30
Best Local Similarity: 24.83% Mismatches: 120
Query Match: 6.65% Indels: 77
DB: Gaps: 14

US-10-018-706-2 (1-322) x US-09-071-035-221 (1-867)
QY 20 LeuIlePheGlyValIleThrThrCysIleLeuAlaGlyCysAlaSerIleProThrTyr 39
DB 31 ATTATTATTGCTTATTCAGTACGACGAGCCCTTTAGCAGGGGAGGAGGAGTTCCTTAT 90
QY 40 AenSerThrSerGlySerGlySerHisArgThrSerGlySerGlyGlyLeuAlaIleGly 59
DB 91 GCGCAAGAAATCGAAGAAATCTTGCTGTAACACAGGAGT----- 132
QY 60 SerGlnValIleThrAspSerGlnGlyValProGln-ArgTyr-----GlnValIleGly 77
DB 133 -----GTTTACACGATGACGATGATACCACTGACCAATGACCAAGTACGACCA 186
QY 77 nGlyAspThrValSerIleAlaGlyArgTyrGlyLeuSerIleProGlnIleGly 97
DB 187 GAGCAACCAACGACGACCAAGTACCAAGTACGAGC-----AACATGGAACCGTCAACA 237
QY 97 sIleAsnAsnLeu----- 237
DB 238 CCAACCGAATCTAGTACGACCTTCAAAACGACGATCTTCTGTTACGACGACGACGAC 297
QY 104 TTYrThrIleTyrThrGlyGlnTyrPleuThrLeuTyrSerIleAspLeuValAlaGly 124
DB 298 GTACCA-----GAGGCCAACA 315
QY 124 uArgSerIleSerSer-GlyValSerThrAlaHis-ThrProSerProValAlaValGln 143
DB 316 ACGCCAAGTACGACGACCAACCAACGACGACCAAGT-----GTACGACGACCA 372
QY 144 SerSerArgProProValGlnGlnHisProAla---ValGlnIleProThrProProVal 162
```

```

DB 373 CCACAGAGCCCAAGTGTACCAAGAAAACCAAGTACCAAGTAAACCAAGCCCA----- 429
QY 163 ValValValIleValIlePheProThrProThrProProValValGlnGlnProValAla 182
DB 430 -----GAAAGCTGTGCTGCAAGTGTCT-----GAAACCAAGTGTGA 468
QY 183 ProProValThrGlnAlaPro-----PheAlaThr 192
DB 469 CCAGAACCAACGACGACCAACCAAGATGTGTGTGTAAGCAATGAGGAATTCACCA 528
QY 193 GlySerSerGlyValMetGlnPheArgTyrProValGlyAlaThrAsn---ProValVal 211
DB 529 GAGGATCT---ACACACAGCCCAAGTCTTCAATGTGAACGAATTAACCTTTCAGAAATA 585
QY 212 ArgArgPheGlyThrThrAlaThrValAlaGlySerThrValThrSerAsnGlyMetTyrPhe 231
DB 586 ACACATGTCCTCCCACTGTGACGACACCCGATTCGAACACCAAGC----- 627
QY 232 SerGlyArgAspGlyAspLeuIleAsnAlaSerAsnAlaGlyThrValIleGlnAlaAsp 251
DB 628 -----GAGAAAGCAATGTGCGAGTGTGAAGGC-----GTTCTTTAACA 669
QY 252 HisAsnMetAspGlyAlaSerIleValIleGlnHisThrAsnGlyPheValSerSerTyr 271
DB 670 CAACGGCTGATGATTGATTAACCAACGATTAACCAATTAATTAATTAACCAAGTGGCAAT 729
QY 272 IleHisIleIleAspAlaGlnValIleThrGlyAspThrValArgThrGlyGlnArgIle 291
DB 730 GTACAAAGTGAAGAAAGTGTGACGGAAGAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 789

RESULT 28
US-09-134-001C-1165
; Sequence 1165, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 1165
; LENGTH: 993
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1165

Alignment Scores:
Pred. No.: 0.112 Length: 993
Score: 111.00 Matches: 36
Percent Similarity: 45.69% Conservative: 17
Best Local Similarity: 31.03% Mismatches: 41
Query Match: 6.65% Indels: 22
DB: Gaps: 4

US-10-018-706-2 (1-322) x US-09-134-001C-1165 (1-993)
QY 15 IleYArgIleGlyLeuIlePheGlyValIleThrThrCysIleLeuAlaGlyCysAla 34
DB 160 ATTGCTAAATTAATCACTTAATGATGATTCACATTAATTAATTCCTTAATCAAGTA 219
QY 35 SerIlePProThrTyrAsnSerThrSerGlySerGlySerHisArgThrSerGlySerGly 54
DB 220 TTGAAA-----GTATCAGGCTCATCTTCAAGGCAACGACCAATGAT 264
QY 55 GlyLeuAlaIleGlySerGlnValIleThrAspSerGlnGlyValProAsnArgTyrGln 74
```

Db 265 GGACAGCTT-----TATACA 279
 Qy 75 VallysgInglyAepThrValSerlyslleaglInarglyrGlyLeuAsnTrpArglu 94
 Db 280 GTTAAAGCTGAATTCATCTCTATGCTGCAAAATAGGTAACAATTATCAAAA 339
 Qy 95 lIeGlyHslleAsnleuAsnSerSerlyrThrIleYrThrGlyInTrpleuThr 114
 Db 340 ATCAGCAACTTAATGGGTAAAT--AACTATCTTATTTCCCGGACAAAGTTGAAA 396
 Qy 115 LeutpSerGlyAepLeuLyValArgluArgSerIleSerSerGly 130
 Db 397 GTT--TCTGTAAAGCAGCAGTTCACGTCGCAAAAGCTAGTGGG 441

RESULT 29

US-08-217-327-7
 Sequence 7, Application US/08217327
 Patent No. 5474925
 GENERAL INFORMATION:
 APPLICANT: John, Maliyakal E
 APPLICANT: Barton, Kenneth A
 TITLE OF INVENTION: Immobilized Proteins in Cotton Fiber
 NUMBER OF SEQUENCES: 16
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Quarles and Brady
 STREET: P.O. Box 2113
 CITY: Madison
 STATE: WI
 COUNTRY: USA
 ZIP: 53701-2113
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/217,327
 FILING DATE:
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/812,233
 FILING DATE: 19-DEC-1991
 ATTORNEY/AGENT INFORMATION:
 NAME: Seay, Nicholas J
 REGISTRATION NUMBER: 27,386
 REFERENCE/DOCKET NUMBER: 1122990831
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 608-251-5000
 TELEFAX: 608-251-9166
 INFORMATION FOR SEQ ID NO: 7:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2077 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 IMMEDIATE SOURCE:
 CLONE: xylanase gene XYNB
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 299..2077
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: 299..632
 OTHER INFORMATION: /function= "cellulose binding
 OTHER INFORMATION: domain"
 US-08-217-327-7

Alignment Scores: 0.394 Length: 2077
 Pred. No.: 110.50 Matches: 74
 Score:

Percent Similarity: 34.49%
 Best Local Similarity: 21.45%
 Query Match: 6.62%
 DB: 1
 Gaps: 16

US-10-018-706-2 (1-322) x US-08-217-327-7 (1-2077)

Qy 1 MetThrValThrIleAlaIleAsnSerGlnAsnGlnPro-----IleYs 16
 Db 287 ATGAGATTAACATGACATTTCCGCAAGCATTAACCGTACCCGGGTAATTTTAA 346
 Qy 17 ArgLeuGly-----LeuIlePheGlyValIleThrThrCysIleLeuAlaGlyCysAla 34
 Db 347 CGTACAACAGCGATTGTGGTGGTGGCAGCATGCACTTAACAGCCCTGGCTTAAGCATCG 406
 Qy 35 SerlySerProThrlyrYrAsnSerThrSerGlySerGlySerHlaThrSerGlySerGly 54
 Db 407 GCTCCCTGTACCTACCATCGATGAC-----GAATGCTCCACC 445
 Qy 55 GlyLeuAlaIleGlySerGlnValIleThrAspSerGlnGlyValProAsnArgTrpGln 74
 Db 446 GGCTTTACCGCAATATACCCCTCAAAACGATACCGGTGCCCATCATTAATCGAAC 505
 Qy 75 VallysgInglyAepThrValSerlyslleaglInarglyrGlyLeuAsnTrpArglu 94
 Db 506 GTGAATTGGCAATCTCAGCAATCGCATGACAGCGGGCTGGATGCAACTTCCGGC 565
 Qy 95 lIeGlyHslleAsnleuAsnSerSerlyrThrIleYrThrGlyInTrpleuThr 114
 Db 566 ACCAACCCCTTACCAACGCCACCAACATGAC----- 595
 Qy 115 LeutpSerGlyAepLeuLyValArgluArgSerIleSer 128
 Db 596 ---TGAACGGCAGACATC---GCCCGAGACATTCATCTCTCGGCTTCAGGCGCAA 649
 Qy 129 ---SerGlyValAsnThr-AlaHisThrProSerProValAlaValGlnSerSerArgPr 147
 Db 650 AAAAATGGCAGCAGCGCGAGCGACCAACCGTACCGGCGCGCTTGAACAGGCAACC 709
 Qy 147 OProValGlnGlnHisProAlaValGlnLyProThrProProValValVal 165
 Db 710 ACCAGCTCTGTGGCTTCAGCTTCAACCCCA---CCACCATTCATCTTCGATCC 766
 Qy 166 -----LysLyPro----- 168
 Db 767 AGTGTGCTCCGCGATGCTGTTGCAAGAACACAGCGGTTTCTGCGCTGGAGTGC 826
 Qy 169 -----ThrProThrProProValValGlnGlnProAlaProValAlaPr 183
 Db 827 ACCATGATTAATACACACCGGCTTACCGGAGTGGCTTGGCAACACCAACAGTCC 886
 Qy 183 OProValThrGlnAlaProPheAla-----ThrGlySerSerGly-----ValMetG1 199
 Db 887 CAGGGG-GCAGCGGTAGTCTGGCGGATAGATGTACACAGCAGTGGCGTCGACCTGCAC 945
 Qy 199 nPheArgTrpProValGlyAlaThrAsnProValValArgArgPheGlyThrAlaThrVa 219
 Db 946 TATCGGCTATGCCAATGTGGAAC-----GCCAATGCAATGCTCATCTGTGAT 996
 Qy 219 lAlaGlySer-----ThrValThrSerAsnGlyMetTrpPhe 232
 Db 997 TAAAGGGGAGCAGCAAGGTAACTATACGTGAGATTGCCACAGCCGGCGCTGGACAC 1056
 Qy 232 rGlyArg-----AspGlyAepLeuIleAsnAlaSerAsn-----Al 244
 Db 1057 CTGGCAAAACCGCACTATGATGTGGTGTGACAGCGCAATATATTTGTCAGTTGTC 1116
 Qy 244 aglyThrValIleGlnAlaAspHisAsnMetArgGlyAlaSerIleValIleGlnHisTh 264
 Db 1117 TGCACGACAGCGGAGGCTTACCCCAATATGATTCGTTAAGTGTGTG----- 1165
 Qy 264 rAsnGlyPheValSerSerlyrIleHisIleYsAspAlaGlnVallyThrGlyAepThr 284

Db 1166 -----GGTGGTAC 1173

QY 284 TValArgThrGly 288

Db 1174 GGTACGGGGCGGT 1186

RESULT 30

US-09-136-574A-2

Sequence 2, Application US/09136574A

Patent No. 6294366

GENERAL INFORMATION:

APPLICANT: Farrington, Graham K.

Anderson, Paige

Gibbs, Moreland

Bergquist, Peter

Daniels, Roy

Morgan, Hugh W.

Williams, Diane P.

TITLE OF INVENTION: Compositions and Methods for Treating Cellulose Containing Fabrics Using Truncated Cellulase Enzyme Compositions

NUMBER OF SEQUENCES: 49

CORRESPONDENCE ADDRESS:

STREET: Spring House and Howson

CITY: Spring House

STATE: PA

COUNTRY: USA

ZIP: 19447

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

OPERATING SYSTEM: DOS

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/136,574A

FILING DATE: 19-Aug-1998

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/932,571

FILING DATE: September 19, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Bak, Mary E.

REGISTRATION NUMBER: 31,215

REFERENCE/DOCKET NUMBER: 1997US001/CIP

TELECOMMUNICATION INFORMATION:

TELEPHONE: 215-540-9200

TELEFAX: 215-540-5818

TELEX: <Unknown>

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 6416 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 2:

US-09-136-574A-2

Alignment Scores:

Score: 2.26 Length: 6416

Percent Similarity: 110.50 Matches: 74

Best Local Similarity: 33.96% Conservative: 34

Query Match: 23.27% Mismatches: 99

Indels: 111

Gaps: 18

US-10-018-706-2 (1-322) x US-09-136-574A-2 (1-6416)

QY 51 SerGlySerGlyGlyLeuAlaIleGlySer-----Gln 61

Db 4003 AGGTGAGCGGTGTTTGGACAGCGGCGACAGACAGAGATAGGCTTAGCATACAG 4062

QY 62 ValIleThrAspSerGlyValProAsnArgTyrGlnValIleGlyValAspThrVal 81

Db 4063 AAGGCACTGGCAGCTACATCAATGATCAATTCGTGAAGAGT----- 4110

QY 82 SerIlySileAlaGlnArgTyrGlyLeuAsnTrpArgIleGlyHisIleAsnAsnLeu 101

Db 4111 -----GGACAGCGCTATATAGAGAACAGAAAGTAAACAGGCTATATAGATGT--- 4158

QY 102 AsnSerSerTyrThrIleTyrThrGlyValInTrpLeuThrLeuTPSerGlyAspLeuIys 121

Db 4159 -----CGATAGTGTGGGA----- 4173

QY 122 ValArgIuArgSerIleSerSerGlyValAsnThrAlaHis-----ThpProSerPro 139

Db 4174 -----AGAGAGCCGAGCAGGGGTACAAAGCCGGGAGGAGAGTACACCGACACCG 4224

QY 140 ValAlaValGlnSerSerArgProProValGlnInHisProAlaValGlnIlyProThr 159

Db 4225 GCACCGAGCCGACATCGACCA-----ACACCAACA 4257

QY 160 ProProValIleValIleValIlyProThrProThrProProValIleGlnInProAla 179

Db 4258 CTACACACCA-----CCGACACCGACACCGACTGTGACGGTGAACCCCACT 4305

QY 180 ProValAlaProProValThrGln----- 187

Db 4306 CCT--ACACCTGGGGTAAACCCCGATGTAAATATCGATCGATCGCCAGGGAGAA 4362

QY 188 -----AlaProPheAlaThrGlySerSerGlyValMetGlnPheArgTyrProVal 204

Db 4363 ACAGAGATTAACCCCGTATATTATGAGCAAAATCGAGATATCCAG----- 4407

QY 205 GlyAlaThrAsnProValIleValArgArgPheGlyThrAlaThrValAlaGlySerThrVal 224

Db 4408 GGTGTGTTCACCT--GCAGAGCAGCTTGT-----GGACACAGATTG 4449

QY 225 ThrSerArgIlyMetTrpPheSerGlyArgAspIlyAspLeuIleAsnAlaSerAspAla 244

Db 4450 ACGGTTTACAT--TGGAGAAC-----AATATGTCCAATGCA 4485

QY 245 GlyThr-----ValIleGln 249

Db 4486 GGGAGTGCCTGTATCATTCAGCGATGATTAATGTTATATTAATGAGTATTAACAGGG 4515

QY 250 AlaAspHisAsnMetAspGlyAlaSerIleValIleGlnHis-----Thr 264

Db 4546 AATGATTAAGAACCTTCAGACAGCTGTGTGAAGCAATTCACAGCGCATTAATTAAGCA 4605

QY 265 AsnGlyPheValSerSerTyrIleHisIleIleAspAlaGlnValIlyThrGlyAsp--- 283

Db 4606 AATGCATATTCAGCATCACTTACAGATGTAGTGTATGTGCGAAGGATGGGAATGT 4665

QY 284 ThrValArgThrGlyGln-----ArgIleAlaSerMetIlyAsnGlnPro 298

Db 4666 ACAGTGAAGCGAGTCAAGACAGCTCCGTCGCGAGAGTGGCTTAAGCTTAATAAAAA 4725

QY 299 SerGlyAlaIleAsnLeuPheGlnPheArgIleSerArgAsnGlyValIlyTyrValAsp 316

Db 4726 GATGTGTCACTGTATTCAGCGCTGACGTGAATGAATGAATATATATATGAT 4779

RESULT 31

US-09-136-574A-46

Sequence 46, Application US/09136574A

Patent No. 6294366

GENERAL INFORMATION:

APPLICANT: Farrington, Graham K.

Anderson, Paige

Gibbs, Moreland

Bergquist, Peter

Daniels, Roy

Morgan, Hugh W.

Williams, Diane P.

TITLE OF INVENTION: Compositions and Methods for Treating Cellulose Containing Fabrics Using Truncated Cellulase Enzyme Compositions

NUMBER OF SEQUENCES: 49
CORRESPONDENCE ADDRESS:
ADDRESSEE: Howson and Howson
STREET: Spring House Corporate Center, P.O. Box 457
CITY: Spring House
STATE: PA
COUNTRY: USA
ZIP: 19477
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/136,574A
FILING DATE: 19-Aug-1998
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/932,571
FILING DATE: September 19, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E.
REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: 1997US001/CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-540-9200
TELEFAX: 215-540-5818
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 46:
SEQUENCE CHARACTERISTICS:
LENGTH: 2029 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 46:
US-09-136-574A-46
Alignment Scores:
Pred. No.: 0.475 Length: 2029
Score: 109.50 Matches: 58
Percent Similarity: 33.47% Conservative: 24
Best Local Similarity: 23.67% Mismatches: 80
Query Match: 6.56% Indels: 83
DB: 4 Gaps: 13
US-10-018-706-2 (1-322) x US-09-136-574A-46 (1-2029)
Qy 34 AlaserlypProthrTyraSn-----SerThrSerglySerglySerrHiaArg 49
Db 193 GCAAGCATGTGACATTCATTTTGTGAAGCTTAGCAGCGAGTGAGCGGATTTAT 252
Qy 50 -----ThrSerglySerglySerglySerrHiaArg 63
Db 253 TACCTGAGTAGGATTGACAGTGGAGCTGG----- 285
Qy 64 ThrAspSergInglyValProAsnArgTyrgInvallySgInglyAspThrValSerys 83
Db 286 -----CAgTTGcAGcCTGgTAAGAGACAGCGGAT 315
Qy 84 lleaaglnArgTyrglyLeuasn-----TryArg 93
Db 316 ATACAGGTAAAGTTTAAACAGAAATGACGACCAATTACATCAGCAGCAGCTGGTCA 375
Qy 94 GluilegIyHsileAsnAsnleu-----AsnSerrTyrrHrIleTyrrHrIle 111
Db 376 TGGTTGACAGACATGACAAATTATGAGAGAAAGTGAAGCGCTGTATGTA---GAT 432
Qy 112 TrpleuThrleuTrpSerglyAspleuIySValArgGluArgSerrIleSerrSerglyVal 131
Db 433 GGTGTTCGTATGCGGG-----CAGAGCCCGGAGAGCGGTG 471
Qy 132 AsnThrIalHsThrProSerProValAlaValGlnSerrArgProProValGlnGln 151
Db 472 ACCCAACTTCTACACCCACACCGGTTTCATCANTCAGCT----- 510

Qy 152 HisProAlaValGlnIySProThrProProValValValIySlySProThrProThr 171
Db 511 -----CTTACACCAACAGCAAGC-----CCACACCTTACA 540
Qy 172 ProProValValGlnGlnProAlaProValAlaProProValThrGluAlaPro----- 189
Db 541 CCTTCTATCACGATTAACACCGAGCGCAACAGCAACCC---ACTCGAGCTCTTGTGTC 597
Qy 190 -----PheAlaThrGlySerrSerglyValMetGlnPhe 200
Db 598 ACAGATATCAAAATGATGATTGTTATTTGCGAGGTAACAAATATGACACAGAT 657
Qy 201 ArgTyProValGlyIatThrAsnProValValAlaArgSphGlyThrAlaThrValAla 220
Db 658 GCGAAGCTGTATGCTTAACAGCA-----CTTATGCTTTGATTTAATACA----- 705
Qy 221 GlySerThrValThrSerrAnglyMetTrpPheSergIyArgAspGlyAspleuIleAsn 240
Db 706 ---GGAACGATGCTTGTATGCTGTGTGAGGTTGTAATCTTAAAGTGATTCGTGAG 762
Qy 241 AlaserAsnAlaGly 245
Db 763 ATTGCAACACAGAGA 777
RESULT 32
US-08-917-320-18
Sequence 18, Application US/08917320
Patent No. 5824508
GENERAL INFORMATION:
APPLICANT: Spaete, Richard and Jackman, Winthrop, T.
TITLE OF INVENTION: No. 5824508 Splicing Variants of gp350/220
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooley Godward Castro Huddleston & Tatum
STREET: 5 Palo Alto Square
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/917,320
FILING DATE: 25-AUG-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/229,291
FILING DATE: April 18, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Luann Cseri
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: AVIR-003/000US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-843-5163
TELEFAX: 415-857-0663
TELEX: 380816 CooleyPA
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 3833 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1014..3734
US-08-917-320-18
Alignment Scores:

Wed Jul 9 10:01:00 2003

us-10-018-706-2.rn1

Page 24

Pred. No.: 1.42 Length: 3833
Score: 109.00 Matches: 66
Percent Similarity: 33.68% Conservative: 30
Best Local Similarity: 23.16% Mismatches: 116
Query Match: 6.53% Indels: 74
DB: 1 Gaps: 13

US-10-018-706-2 (1-322) x US-08-917-320-18 (1-3833)

```
QY 35 SerLyPProThrThyAsnSerThrseryGlySerHisArgThrseryGly 54
DB 2301 ACCTCCCTTACCTTGAATACATGATTTGCTATCCCAATACACACAGGT----- 2354
QY 55 GtLeuAlaIleGlySerGlnValIleThrsPserGlnGlyValProAsnArgTyr--- 73
DB 2355 -----CTACCCAGCTTACTCAAGTGCCTTACCAACCTCACC 2390
QY 74 ---GlnValLysGlnGlyAspThrValSerLysIleAlaGlnArgTyrGlyLeuAsnTrp 92
DB 2391 GCACCTGCACAGCAGGCCCTGTATCC----- 2420
QY 93 ArgGlnIleGlyHisIleAsnAsnLeuAsnSerSerTyrTrpIleTyrTrpGlyGlnTrp 112
DB 2421 -----ACCGCGATGTATCACCACCAACACCGCGACAGTACGAGCGCATCACC 2474
QY 113 LeuThr-----LeuTrpSerGlyAspLeuLysValArgGlnArgSerIleSer 128
DB 2475 GTGACACCAAGTCATCTCCATGGACACGACAGAGTAAGTAGGCCCGACATGACC 2534
QY 129 SerGlyValAsnThrAlaHisThrPserProValAlaValGlnSerSerArgProPro 148
DB 2535 AGCTCCACCTCAGCAGTACTACCCCAACCCCAATGTCACC---ACCCACCCACGAC 2591
QY 149 Val-----GlnGlnHisProAlaValGlnLysProThrPro 160
DB 2592 GTGACTACCCCAACCCCAATGTCACAGCCGCCACAGTACTACCCACCCCA 2651
QY 161 -----ProValValValLysLysProThr 169
DB 2652 AATGCCACAGCCCTTGGGAAACAAAGTCTTACCTCAGACAGTACTACCCCA 2711
QY 170 Pro-----ThrProValValGlnGlnProAlaProValAlaProProValThrGln 187
DB 2712 CCAATGCCACAGCCCTTGGGAAACAAAGTCTTACCTCAGACAGTACTACCCCA 2771
QY 188 AlaProPheAlaThrGlySerSerGlyValMetGlnPheArgTyrProValGlyAlaThr 207
DB 2772 ACCCCAAATGCCACAGCCCTTGGGAAACAAAGTCTTACCTCAGACAGTACTACCCCA 2807
QY 208 AsnProValValArgArgPheGlyThrAlaThrValAlaGlySerThrValThrSerAsn 227
DB 2808 AGCCCTTGGGAAACAAAGTCTTACCTCAGACAGTACTACCCCAATGCCCAATGCCCA 2849
QY 228 GlyMetTrpPheSerGlyValArgPheGlyAspLeuIleAsnAlaSerAsn-----Ala 244
DB 2850 GGCCCTTACCTTGGGAAACAAAGTCTTACCTCAGACAGTACTACCCCAATGCCCAATGCCCA 2906
QY 245 GlyThr-----ValIleGlnAlaAspHisAsnMetAspGlyValSerIle 259
DB 2907 GGACACAGTCCACAGCCCTTGGGAAACAAAGTCTTACCTCAGACAGTACTACCCCAATGCCCAATGCCCA 2966
QY 260 ValIleGlnHisThrAsnGlyPheValSerSerTyrIleHisIleLysAspAlaGlnVal 279
DB 2967 ACAGCCCAACATTA-CATTAATTTCAAGTTCAACCTTCTCAGTCACTGACAGACCACTTC 3025
QY 280 LysThrGlyAspThr 284
DB 3026 AAACCCAGAGCACT 3040
```

RESULT 33
PCT-US95-04611A-18
Sequence 18, Application PC/TUS9504611A
GENERAL INFORMATION:

APPLICANT: Spaete, Richard and Jackman, Winthrop T.
TITLE OF INVENTION: Non Splicing Variants of gp350/220
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooley Godward Castro Huddleson & Tatum
STREET: 5 Palo Alto Square
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patient In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/04611A
FILING DATE:
CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 08/229,291
FILING DATE: April 18, 1994
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Luanm Geert
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: AVIR-003/00US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-857-0663
TELEFAX: 415-843-5163
TELEX: 380816 COOLEYPA
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 3833 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1014..3734
PCT-US95-04611A-18

Alignment Scores:

Pred. No.: 1.42 Length: 3833
Score: 109.00 Matches: 66
Percent Similarity: 33.68% Conservative: 30
Best Local Similarity: 23.16% Mismatches: 116
Query Match: 6.53% Indels: 74
DB: 5 Gaps: 13

US-10-018-706-2 (1-322) x PCT-US95-04611A-18 (1-3833)

```
QY 35 SerLyPProThrThyAsnSerThrseryGlySerHisArgThrseryGly 54
DB 2301 ACCTCCCTTACCTTGAATACATGATTTGCTATCCCAATACACACAGGT----- 2354
QY 55 GtLeuAlaIleGlySerGlnValIleThrsPserGlnGlyValProAsnArgTyr--- 73
DB 2355 -----CTACCCAGCTTACTCAAGTGCCTTACCAACCTCACC 2390
QY 74 ---GlnValLysGlnGlyAspThrValSerLysIleAlaGlnArgTyrGlyLeuAsnTrp 92
DB 2391 GCACCTGCACAGCAGGCCCTGTATCC----- 2420
QY 93 ArgGlnIleGlyHisIleAsnAsnLeuAsnSerSerTyrTrpIleTyrTrpGlyGlnTrp 112
DB 2421 -----ACCGCGATGTATCACCACCAACACCGCGACAGTACGAGCGCATCACC 2474
QY 113 LeuThr-----LeuTrpSerGlyAspLeuLysValArgGlnArgSerIleSer 128
DB 2475 GTGACACCAAGTCATCTCCATGGACACGACAGTAAGTACCCCAATGACC 2534
```

```

Qy 129 SerGlyValAsnThrAlaHisThrProSerProValAlaValGlnSerSerArgProPro 148
Db 2535 AGCTCCACTCAGCAGTACTATACCCCAACCCCAATGCGACC--AGCCCCACCCCAACA 2591
Qy 149 Val-----GlnGlnHisProAlaValGlnLysProThrPro 160
Db 2592 GTGACTACCCCAACCCCAATGCGACAGCCCAACCCCAAGTACTATACCCCAACCCCA 2651
Qy 161 -----ProValValValValLysProThrPro 169
Db 2652 AATGCCACAGCCCACTTGGAAGAAAACAAGTCTTACTCTAGCAGTACTATACCCCAACC 2711
Qy 170 Pro-----ThrProProValValGlnGlnProAlaProValAlaProProValThrGlu 187
Db 2712 CCAATATGCCACAGCCCACTTGGAAGAAAACAAGTCTTACTCTAGCAGTACTATACCCCA 2771
Qy 188 AlaProPheAlaThrGlySerSerGlyValMetGlnPheArgTyrProValGlyAlaThr 207
Db 2772 ACCCCCAATGCCACAGCCCACTTGGAAGAAAACAAGTCTTACTCTAGCAGTACTATAC 2807
Qy 208 AsnProValValArgArgPheGlyThrAlaThrValAlaGlySerThrValThrSerAsn 227
Db 2808 AGCCCC-----ACCTCAGCAGTACTATACCCCAACCCCAATGCGACC 2849
Qy 228 GlyMetTrpPheSerGlyArgAspGlyAspLeuIleAsnAlaSerAsn-----Ala 244
Db 2850 GGCCCTACTGTG--GGAGAAACAAGTCCACAGGCAATATGCCACCAACACACCTTAGA 2906
Qy 245 GlyThr-----ValIleGlnAlaAspHisAsnMetAspGlyAlaSerIle 259
Db 2907 GGAACCAATGCCACCCAGTACTATACCCCAACCAACCAACCAACCAAGTCTTACC 2966
Qy 260 ValIleGlnHisThrAsnGlyPheValSerSerTyrIleHisIleAspAlaGlnVal 279
Db 2967 ACAGGCCCAACTTA--CATTACTTCAAGTTCACCTTCCACAGTACTAGACCCCACTTC 3025
Qy 280 LysThrGlyAspThr 284
Db 3026 AAACCCAGAGACT 3040

```

```

RESULT 34
US-08-783-774-1
Sequence 1, Application US/08783774
Patent No. 6054130
GENERAL INFORMATION:
APPLICANT: Spaete, Richard
TITLE OF INVENTION: NON-SPLICING VARIANTS OF
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESS: Pennie & Edwards
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036/2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
OPERATING SYSTEM: DOS
SOFTWARE: FASTA version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/783,774
FILING DATE: 15-JAN-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7682-037
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-790-9090
TELEFAX: 212-869-8864

```

```

TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5931 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 1014...3734
OTHER INFORMATION:
US-08-783-774-1
Alignment Scores:
Pred. No.: 2.79 Length: 5931
Score: 109.00 Matches: 66
Percent Similarity: 33.68% Conservative: 30
Best Local Similarity: 23.16% Mismatches: 116
Query Match: 6.53% Indels: 74
DB: 3 Gaps: 13
US-10-018-706-2 (1-322) x US-08-783-774-1 (1-5931)
Qy 35 SerLysProThrTyrAsnSerThrSerGlySerGlySerHisArgThrSerGlySerGly 54
Db 2301 ACCCTCCCTACTTGAATATACAGTATTGCTATCCCAATACAGCAGGT----- 2354
Qy 55 GlyLeuAlaIleGlySerGlnValIleThrAspSerGlnGlyValProAsnArgTyr-- 73
Db 2355 -----CTACCAAGCTCTACTACAGTCTGCTTACCAACCTCACC 2390
Qy 74 ---GlnValLysGlnGlyAspThrValSerLysIleAlaGlnArgTyrGlyLeuAsnTrp 92
Db 2391 GCACCTGCAGACAGCAGGCCCTCAGTATCC----- 2420
Qy 93 ArgGluIleGlyHisIleAsnLeuAsnSerSerTyrThrIleTyrThrGlyGlnTrp 112
Db 2421 -----ACCGCGATGTCCACAGCCCAACAGCAGCCGACACAGTCAAGCGCATCACC 2474
Qy 113 LeuThr-----LeuTrpSerGlyAspLeuLysValArgGluArgSerIleSer 128
Db 2475 GTGACACCAAGTCCATCTCCATGAGCAAGGACAGCAAGAAAGTAAAGGCCCGCATGAC 2534
Qy 129 SerGlyValAsnThrAlaHisThrProSerProValAlaValGlnSerSerArgProPro 148
Db 2535 AGCTCCACTCAGCAGTACTATACCCCAACCCCAATGCGACC--AGCCCCACCCCAACA 2591
Qy 149 Val-----GlnGlnHisProAlaValGlnLysProThrPro 160
Db 2592 GTGACTACCCCAACCCCAATGCGACAGCCCAACCCCAAGTACTATACCCCAACCCCA 2651
Qy 161 -----ProValValValValLysProThrPro 169
Db 2652 AATGCCACAGCCCACTTGGAAGAAAACAAGTCTTACTCTAGCAGTACTATACCCCAACC 2711
Qy 170 Pro-----ThrProProValValGlnGlnProAlaProValAlaProProValThrGlu 187
Db 2712 CCAATATGCCACAGCCCACTTGGAAGAAAACAAGTCTTACTCTAGCAGTACTATACCCCA 2771
Qy 188 AlaProPheAlaThrGlySerSerGlyValMetGlnPheArgTyrProValGlyAlaThr 207
Db 2772 ACCCCCAATGCCACAGCCCACTTGGAAGAAAACAAGTCTTACTCTAGCAGTACTATAC 2807
Qy 208 AsnProValValArgArgPheGlyThrAlaThrValAlaGlySerThrValThrSerAsn 227
Db 2808 AGCCCC-----ACCTCAGCAGTACTATACCCCAACCCCAATGCGACC 2849
Qy 228 GlyMetTrpPheSerGlyArgAspGlyAspLeuIleAsnAlaSerAsn-----Ala 244
Db 2850 GGCCCTACTGTG--GGAGAAACAAGTCCACAGGCAATATGCCACCAACACCTTAGA 2906
Qy 245 GlyThr-----ValIleGlnAlaAspHisAsnMetAspGlyAlaSerIle 259

```

Db 2907 GGAACACATCCACCCAGTAGTTACCGACCAACAAAATGCAACAGTGTCTTACC 2966
Qy 260 ValIIGlnHsthranglyPheValSerSerTyrlleHisIlelyaspAlaGlnVal 279
Db 2967 ACAGGCCAACATA-CATAACTTCAGATTCACTCTTCATGTCTAGTACAGACCAGTTC 3025
Qy 280 LysThrGlyAspThr 284
Db 3026 AAACCCAGAGACT 3040
RESULT 35
US-09-556-706B-1
Sequence 1, Application US/09556706B
Patent No. 6458364
GENERAL INFORMATION:
APPLICANT: Spaete, Richard
APPLICANT: Jackman, Winthrop
TITLE OF INVENTION: NON SPLICING VARIANTS OF GP350/220
FILE REFERENCE: 7682-050-999
CURRENT APPLICATION NUMBER: US/09/556,706B
CURRENT FILING DATE: 2000-04-24
PRIOR FILING DATE: 1997-01-15
PRIOR APPLICATION NUMBER: 08/783,774
PRIOR FILING DATE: 1994-04-18
NUMBER OF SEQ ID NOS: 19
SOFTWARE: PatentIn version 3.0
SEQ ID NO 1
LENGTH: 5931
TYPE: DNA
ORGANISM: Virus
FEATURES:
OTHER INFORMATION: GP350/220
US-09-556-706B-1
Alignment Scores:
Pred. No.: 2.79 Length: 5931
Score: 109.00 Matches: 66
Percent Similarity: 33.68% Conservative: 30
Best Local Similarity: 23.16% Mismatches: 116
Query Match: 6.53% Indels: 74
DB: 4 Gaps: 13
US-10-018-706-2 (1-322) x US-09-556-706B-1 (1-5931)
Qy 35 SerLyPProThrTyraSerThrSerGlySerGlySerHisArgThrSerGlySerGly 54
Db 2301 ACCTCCCTACCTTGAATACCACTGATTGCTGATCCCAATACACGACAGT----- 2354
Qy 55 GlyLeuAlaIleGlySerGlnValIleThrAspSerGlnGlyValProAsnArgTyrr--- 73
Db 2355 -----CTACCCAGCTCTACTACAGTCTGCTTACCAACCTCACC 2390
Qy 74 -----GlnValIlyGlnGlyAspThrValSerLyIleAlaGlnArgTyrrGlyLeuAsnTrp 92
Db 2391 GCACTCGACAGACAGAGCCCACTGATTC----- 2420
Qy 93 ArgGluIleGlyHisIleAsnLeuAsnSerSerTyrlleTyrrIleTyrrGlnTrp 112
Db 2421 -----ACCGCGGATGTCCAGCCCAACAGCCGCGACAGTCAAGCCGATCACC 2474
Qy 113 LeuThr-----LeuTrpSerGlyAspLeuIlyValArgGlnArgSerIleSer 128
Db 2475 GTGACACCAAGTCCATCTCCATGGACAAAGGACGAAAGTAAAGGCCCCGACATGACC 2534
Qy 129 SerGlyValAsnThrAlaHisThrProSerProValAlaValGlnSerSerArgProPro 148
Db 2535 AACTCCACCTCACAGTGAATCCCAACCCCAAAATGCCACC-----AGCCCCACCCACAGA 2591
Qy 149 ValI-----GlnGlnHisProAlaValGlnIlyProThrPro 160
Db 2592 GTGACTACCCCAACCCCAATGCGACAGCCCAACCCCAAGTGAATACCCCAACCCCA 2651

Qy 161 -----ProValValValValIlyAspProThr 169
Db 2652 AATGCCACAGCCCACTTGGAAMAAAGTCTTACCTGACAGTACTACCCCAAC 2711
Qy 170 Pro-----ThrProValValGlnGlnProAlaProValAlaProProValThrGlu 187
Db 2712 CCAATGACACAGCCCACTTGGAAMAAAGTCTTACCTGACAGTACTACCCCA 2771
Qy 188 AlaProPheAlaThrGlySerSerGlyValMetGlnPheArgTyrrProValIlyAlaThr 207
Db 2772 ACCCAATGACACAGCCCACTTGGAAMAAAGTCTTACCTGACAGTACTACCCCA 2807
Qy 208 AsnProValValArgArgPheGlyThrAlaThrValAlaGlySerThrValThrSerAsn 227
Db 2808 AGCCCC-----ACCTCAGCAGTGAATACCCCAACCCCAATGCCAC 2849
Qy 228 GlyMetTrpPheSerGlyArgAspGlyAspLeuIleAsnAlaSerAsn-----Ala 244
Db 2850 GGCCTTACTGTG-----GGAGAAACAGTCCACAGGCAATGCCACACACACCTTAC 2906
Qy 245 GlyThr-----ValIleGlnAlaAspHisAsnMetAspGlyAlaSerIle 259
Db 2907 GGAACATGCCACCCAGTGTCTTACCGACCAACCAAAAATGCAACCAAGTGTCTTACC 2966
Qy 260 ValIleGlnHsthranglyPheValSerSerTyrlleHisIlelyaspAlaGlnVal 279
Db 2967 ACAGGCCAACATA-CATAACTTCAGATTCACTCTTCATGTCTAGTACAGACCAGTTC 3025
Qy 280 LysThrGlyAspThr 284
Db 3026 AAACCCAGAGACT 3040
RESULT 36
US-08-314-309A-1
Sequence 1, Application US/08314309A
Patent No. 5677141
GENERAL INFORMATION:
APPLICANT: ISOGAMI, TAKAO
APPLICANT: FUKAGAWA, MASAO
APPLICANT: IWAMI, MORITA
APPLICANT: ARAMORI, ICHIRO
APPLICANT: KOJO, HITOSHI
TITLE OF INVENTION: PROCESS FOR PRODUCING 7-AMINOCEPHEM
TITLE OF INVENTION: COMPOUND OR SALTS THEREOF
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT,
STREET: 1755 S. Jefferson Davis Highway, Suite 400
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/314,309A
FILING DATE: 30-SEP-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/631,906
FILING DATE: 21-DEC-1990
ATTORNEY/AGENT INFORMATION:
NAME: Oblon, No. 5677141man F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 18-863-0 CONT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 413-3000
TELEFAX: (703) 413-2220

```

TELEX: 24885 OPAT UR
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3172 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
US-08-314-309A-1

Alignment Scores:
Pred. No.: 1.32 Length: 3172
Score: 108.00 Matches: 67
Percent Similarity: 33.44% Conservative: 36
Best Local Similarity: 21.75% Mismatches: 113
Query Match: 6.47% Indels: 93
DB: 1 Gaps: 14

US-10-018-706-2 (1-322) x US-08-314-309A-1 (1-3172)
QY 39 Tyranser-----ThrserylserylserysHisArgthrsr 51
   |||:||||:
Db 1570 TACACGCCGCTCGCGGAGACACCCGACACCGCGCACGCCGCTGCGTGGC 1629
QY 52 Glyserglyyleualaileglyserglnvaliethrpserserglnlyvalproasn 71
   |||:||||:
Db 1630 ACCATTGGCGGACGAGCTACGCGGCTGCCAAGAACCAACCTCATGCTGTC----- 1683
QY 72 ArgTyrGlnValIysGlnGlyAspThrValSerIysIleAlaGlnArgTyrGlyLeuasn 91
   |||:||||:
Db 1684 -----AAGGCTTCCTCGGGGATCTTCGAGCTCTTCCATCATCATCTTGACGGCTTCAAC 1737
QY 92 TrparGluIleGlyHisIleasn-----AsnLeuasnSerSerThrIleTyr 108
   |||:||||:
Db 1738 TGGCGCCGAAAGATATCATCAACGCGGCGCGCGCAACAGGCTGCATGACATGTC 1797
QY 109 ThrGlyGlnTrpLeuThrLeuTrpSerGlyAspLeuIysValArgGluArgSerIleSer 128
   |||:||||:
Db 1798 CTGGGTGAGCTATACCCCTTTTTCCTCCGTCAC----- 1830
QY 129 SerGlyValAsnThrAlaHisThrProSerProValAlaValGlnSerSerArgProPro 148
   |||:||||:
Db 1831 -----ACCAAGACACTCAATTCCTTGTCTTAACACCACTAAACTCCCCCC 1878
QY 149 ValGln-----GlnHisProAla 154
   |||:||||:
Db 1879 TTGCGAGGTGGCTACTCTTCTGCTTCAACACGCGCTCAACCTGCTTACCTCCGCG 1938
QY 155 ValGlnIysProThr---ProProValValValIysIysProThrProThrPro--- 172
   |||:||||:
Db 1939 GCGTCTCTCGCTGCGCGCGCGGACAGATTAACAGAGCGCGCCCACTACTCCCGCG 1998
QY 173 -----ProValValGlnGlnProAlaProValAlaProProValThrGlnAlaProPhe 190
   |||:||||:
Db 1999 CCTCGCGCGCAACCCCATCAACGCTGCTCCCTCCCAACTGGGCGCGCTCCAGCT 2058
QY 191 AlaThrGlySerSerGlyValMetGlnPheArgTyrProValGlyAlaThrAsnProVal 210
   |||:||||:
Db 2059 TCAGCACTACGCGCTCC-GTGTCTGACATCTTGGCCCC----- 2096
QY 211 ValArgArgPheGlyThrAlaThrValAlaGlySerThrValThrSerAsnGlyMetTyr 230
   |||:||||:
Db 2097 -----GGAACACACATCTCTTCC-----GCCCTG 2120
QY 231 PheSerGlyArgAspGlyAspLeuIleAsnAlaSerAsnAlaGlyThrValIleGlnAla 250
   |||:||||:
Db 2121 ATCGCGCGGCAC-----TCGCGCC 2138
QY 251 AspHisAsnMetAspGlyAlaSerIleValIleGlnHisThrAsnGlyPheValSerSer 270
   |||:||||:
Db 2139 ACCAACACACATCTCGGACCTCTCCATGCGCACCCCACTGTCACCGGCTGCTC---CTC 2195
QY 271 TyrIleHisIleIysAspAlaGlnValIysThrGlyAspThrValArg----- 286

```

```

Db 2196 TACCTCAGCGCCTCGAGGCTGTACACCACTCTGCGCTGCGCGCGCTCAACGCTCTG 2255
QY 287 -----ThrGlyGlnAlaIleAlaSerMetIysAsnGlnProSerGlyAlaAlaLeuPhe 304
   |||:||||:
Db 2256 GCCACACCGGCGCT-----GTCTCAACCTGAGCTCGGTAGC----- 2294
QY 305 GlupheArgIleSerArgAsnGly 312
   |||:||||:
Db 2295 CCCAACCGCATCTCTACACCGC 2318

RESULT 37
US-09-258-371-9
; Sequence 9, Application US/09258371
; Patent No. 5986078
; GENERAL INFORMATION:
; APPLICANT: Gaikavsev, Igor
; APPLICANT: Riadowol, Karl
; TITLE OF INVENTION: DNA SEQUENCE ENCODING THE TUMOR
; TITLE OF INVENTION: SUPPRESSOR GENE INGI
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: 699 Prince Street
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/258,371
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/751,230
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Mooi, Leslie A.
; REGISTRATION NUMBER: 37,047
; REFERENCE/DOCKET NUMBER: 028722-144
; TELEPHONE: 415-854-7400
; TELEFAX: 415-854-8275
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2061 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 16..900
; US-09-258-371-9

Alignment Scores:
Pred. No.: 0.759 Length: 2061
Score: 107.50 Matches: 66
Percent Similarity: 38.03% Conservative: 42
Best Local Similarity: 23.24% Mismatches: 102
Query Match: 6.44% Indels: 74
DB: 2 Gaps: 14

US-10-018-706-2 (1-322) x US-09-258-371-9 (1-2061)
QY 22 PheGlyValIleThrThrCysIleLeuAlaGly-CysAlaSerIysProThrTyrAsn-- 40
   |||:||||:
Db 41 TTCCAGATATAGCAGTAGACGATGATCCGGGCTGTGGCTCGGGCGCGGCTGCAGTT 100

```

Wed Jul 9 10:01:00 2003

us-10-018-706-2.rn1

Page 28

41 -----SerThrSerGlySerGlySerHisArgThrSerGlySer-----G1 54
101 CGAGCCGCTCCCGGAGCCCGGGGGCGGCTCGGACAGATTGAGCCCGCATCTTGC 160
54 YGLYLEUMLALEGYSERGINVALILE-----ThrAspSerGlnGly-ValProAsn 72
161 TGACCCGAGGTGGGCGCGCGCTGGCGGTGAAAACATCCTGAAGAGCTAGACGAGT 220
72 rgytyrgln-----VallysglnGlyAspThrValSerIys-----IleA 85
221 GCTACGAGCGCTTCATCGGACAGACAGCGGGCGGACAGAGATCCAGATCGTAC 280
85 IAGLNAITGTYRGLYLEUANTTPARGIULILEGIIYHIS-----IleAsn 100
281 TGCACCGCGCGCTGATCCGACCGACAGAGCTGGCGGACAGAGATCCAGATCGTAC 340
100 snLeuAsn-SetSerTyThrIleTyThrGlyGlnTTPLeuThrLeuTTPSer----- 117
341 AGATGTGAGTGTGTGAGAAACCGACGCGGAGGTGACACGACGAGCTGTTCG 400
118 -----Gly 118
401 AGCGCAGACAGAGCTGGCGGACAGAGTGGGACAGCGGAGTTGGCGGACAGGCG 460
119 AspleuLysValArgGluArgSerIleSerSerGlyValAsnThrAlaHisThrProSer 138
461 CCAATGGCGATGCGGTAGCGCAGCTCTGACAAACCAAGCAAGAGGCTCACGCGGAGC 520
139 -----ProValAlaValGlnSerSerArgProProVal 149
521 GCAACACAGAGAACCGTGAAGACGCGTCCAGACACAGACAGACAGACGCGGCT--- 577
150 GINGLNHISProAlaValGlnLysProThrProProValValValLysProThr 169
578 CGGGCACACCCAG 628
170 ProThrPro---ProValValGlnGlnProAlaProValAlaProProValThrGluAla 188
629 CCAAGGCCAAGCGGAG 679
189 ProPheAlaThrGlySerSerGlyValMetGlnPheArgTyProValGlyAlaThrAsn 208
680 -----CCAACGAAACCCAGTACTGTCTGTGCAACAGAGTCTCTAGGGAGATGATCG 733
209 ProValValArgArgPheGlyThrAlaThrValAlaGlySerThrValThrSerGlnGly 228
734 GCT-----GCGACACAGCAGAGAGTCCCATCGAGTGGTCCACTCTCTCGT 778
229 MetTPPheSerGlyArgAspGlyAspLeuIleAsnAlaSerAsnAlaGlyThrValIle 248
779 GCGTGTG-----GCTCAATCATATMAACCAAGGAGCAAGTGTACTGTCCCA 823
249 GlnAla 250
824 AGTGCC 829

RESULT 38
US-08-751-230-9
Sequence 9, Application US/08751230
Patent No. 6117633

GENERAL INFORMATION:
APPLICANT: Garkavtsev, Igor
APPLICANT: Ribaowol, Karl
TITLE OF INVENTION: DNA SEQUENCE ENCODING THE TUMOR
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Birn, D. Lane, Swecker & Mathis
STREET: 699 Prince Street
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-1404

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/751,230
FILING DATE: 15-NOV-1996
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/569721
FILING DATE: 08-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Mool, Leslie A.
REGISTRATION NUMBER: 37,047
REFERENCE/DOCKET NUMBER: 028722-144
TELEPHONE: 415-854-7400
TELEFAX: 415-854-8275
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 2061 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 16..900
US-08-751-230-9
Alignment Scores:
Pred. No.: 0.759 Length: 2061
Score: 107.50 Matches: 66
Percent Similarity: 38.03% Conservative: 42
Best Local Similarity: 23.24% Mismatches: 102
Query Match: 6.44% Indels: 74
DB: Gaps: 14
US-10-018-706-2 (1-322) x US-08-751-230-9 (1-2061)
22 PheGlyValIleThrThrCysIleLeuAlaGly-CysAlaSerLysProThrTyraa- 40
41 TTCACAGATATAGCAGATAGCAGATATCCCGGCTGTGGTCCGGGGCGGAGTTCAGTT 100
41 -----SerThrSerGlySerGlySerHisArgThrSerGlySer-----G1 54
101 CGAGCCGCTCCCGGAGCCCGGGGGCGGCTCGGACAGATTGAGCCCGCATCTTGC 160
54 YGLYLEUMLALEGYSERGINVALILE-----ThrAspSerGlnGly-ValProAsn 72
161 TGACCCGAGGTGGGCGCGCGCTGGCGGTGAAAACATCCTGAAGAGCTAGACGAGT 220
72 rgytyrgln-----VallysglnGlyAspThrValSerIys-----IleA 85
221 GCTACGAGCGCTTCATCGGACAGACAGCGGGCGGACAGAGATCCAGATCGTAC 280
85 IAGLNAITGTYRGLYLEUANTTPARGIULILEGIIYHIS-----IleAsn 100
281 TGCACCGCGCGCTGATCCGACCGACAGAGCTGGCGGACAGAGATCCAGATCGTAC 340
100 snLeuAsn-SetSerTyThrIleTyThrGlyGlnTTPLeuThrLeuTTPSer----- 117
341 AGATGTGAGTGTGTGAGAAACCGACGCGGAGGTGACACGACGAGCTGTTCG 400
118 -----Gly 118
401 AGCGCAGACAGAGCTGGCGGACAGAGTGGGACAGCGGAGTTGGCGGACAGGCG 460
119 AspleuLysValArgGluArgSerIleSerSerGlyValAsnThrAlaHisThrProSer 138
461 CCAATGGCGATGCGGTAGCGCAGCTCTGACAAACCAAGCAAGAGGCTCACGCGGAGC 520

QY 139 -----ProValAlaValGlnSerSerArgProProVal 149
Db 521 GCAACAAGAGAACCGTGAGAACGGCTCCAGCAACAGACAGACGCGCT--- 577
QY 150 GlnGlnHisProAlaValGlnLysProThrProProValValValLysLysProThr 169
Db 578 CGGGACACCCAGAGAGAGAGGCGCAAGACTCCA-----AGAAAGAAAGCGCT 628
QY 170 ProThrPro---ProValValGlnGlnProAlaProValAlaProProValThrGluAla 188
Db 629 CCAAGGCGAAGCGGAGCGAGAGCGGCTCCCTGCCAGCTCCCAATCCAGC----- 679
QY 189 ProPheAlaThrGlySerSerGlyValMetGlnPheArgTyrProValGlyAlaThrAsn 208
Db 680 -----CCAACGAAACCCACGACTCTCTGTGCAACCAAGCTCTCTATGCGAGATGATCG 733
QY 209 ProValValArgArgPheGlyThrAlaThrValAlaGlySerThrValThrSerAsnGly 228
Db 734 GCT-----GCCGCAACAGCAGAGTCCCATCGAGTGTTCACATCTCTCGT 778
QY 229 MetTrpPheSerGlyArgAspGlyAspLeuIleAsnAlaSerAsnAlaGlyThrValIle 248
Db 779 GCGTGG-----GGCTCAATCATTAACCAAGGCGCAAGGTACTGTCTCCA 823
QY 249 GlnAla 250
Db 824 AGTGC 829

RESULT 39
US-09-499-082-9
Sequence 9, Application US/09499082
Patent No. 6143522
GENERAL INFORMATION:
APPLICANT: Helbing, Karen C.
APPLICANT: Klabowol, Karl
APPLICANT: Johnston, Randall N.
APPLICANT: Garkavtsev, Igor
TITLE OF INVENTION: METHODS OF MODULATING APOPTOSIS
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: 699 Prince Street
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09499.082
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/828.158
FILING DATE: 27-MAR-1997
APPLICATION NUMBER: US 08/751230
FILING DATE: 15-NOV-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/569721
FILING DATE: 08-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Moori, Leslie A.
REGISTRATION NUMBER: 37,047
REFERENCE/DOCKET NUMBER: 028722-148
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-854-7400
TELEFAX: 650-854-8275
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 2061 base pairs

TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 16..900
US-09-499-082-9

Alignment Scores:
Pred. No.: 0.759 Length: 2061
Score: 107.50 Matches: 66
Percent Similarity: 38.03% Conservative: 42
Best Local Similarity: 23.24% Mismatches: 102
Query Match: 6.44% Indels: 74
Gaps: 14

US-10-018-706-2 (1-322) x US-09-499-082-9 (1-2061)

QY 22 PheGlyValIleThrThrCysIleLeuAlaGly-CysAlaSerLysProThrTyrAsn-- 40
Db 41 TTCCAGATATPAGATGACATGATCCGGGCTGTGCTCGGGCCGGGCTGCAGTT 100
QY 41 -----SerThrSerGlySerGlySerHisArgThrSerGlySer-----G1 54
Db 101 CGAGCCGCTCCCGCGAGCCCGCGCGCGCGCTCGAGACAGTTTCAAGCCGATCTTGC 160
QY 54 YGlyLeuAlaIleGlySerGlnValIle-----ThrAspSerGlnGly-ValProAsn 72
Db 161 TGACCCGAGGGGTGGGGCGCGCGCTGTGGCAACAGATCTTGAAGAGCTAGACGAGT 220
QY 72 rGlyrGln-----ValLysGlnLysPheThrValSerLys-----IleA 85
Db 221 GCTACGAGCGCTTCAATGCGGAGACAGACGGGGCGCAAGCGCGAGTCTGCATGTG 280
QY 85 IaglnArgTyrGlyLeuAsnTrpArgLuiIleGlyHis-----IleAsn 100
Db 281 TGCAGCGCGCGCTGATCCGACCGAGAGCTCGGCGCAGAAAGATCCAGATCGTAGCC 340
QY 100 snLeuAsn-SerSerTyrThrIleTyrThrGlyGlnTrpLeuThrLysPheSer----- 117
Db 341 AGATGTGTGAGCTGTGTGAGAGAACCGCAGCGGAGGTGACACGATGAGCTGTTCG 400
QY 118 -----Gly 118
Db 401 AGCGCGACGAGAGCTGGCGGACACAGTGGGCAACAGCGGAGTTGGCGCGGACAGC 460
QY 119 AspLeuLysValArgGluArgSerIleSerSerGlyValAsnThrAlaHisThrProSer 138
Db 461 CCAATGGCGATCGCGTGGCCGACTGTGACAAAGCCCAACAGACGCTCACGGCGGACG 520
QY 139 -----ProValAlaValGlnSerSerArgProProVal 149
Db 521 GCAACAAGAGAACCGTGAGAACCGCTCCAGCAACAGACAGACGAGCGCGCT--- 577
QY 150 GlnGlnHisProAlaValGlnLysProThrProProValValValLysLysProThr 169
Db 578 CGGGACACCCAGAGAGAGAGGCGCAAGACTCA-----AGAAAGAAAGCGCT 628
QY 170 ProThrPro---ProValValGlnGlnProAlaProValAlaProProValThrGluAla 188
Db 629 CCAAGGCGAAGCGGAGCGAGAGCGCTCCCTGCCAGCTCCCATGAGC----- 679
QY 189 ProPheAlaThrGlySerSerGlyValMetGlnPheArgTyrProValGlyAlaThrAsn 208
Db 680 -----CCAACGAAACCCACGACTCTCTGTGCAACCAAGCTCTCTATGCGAGATGATCG 733
QY 209 ProValValArgArgPheGlyThrAlaThrValAlaGlySerThrValThrSerAsnGly 228
Db 734 GCT-----GCCGCAACAGCAGAGTCCCATCGAGTGTTCACATCTCTCGT 778
QY 229 MetTrpPheSerGlyArgAspGlyAspLeuIleAsnAlaSerAsnAlaGlyThrValIle 248

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 3, 2003, 21:28:33 ; Search time 146 Seconds

(without alignments)
3442.734 Million cell updates/sec

Title: US-10-018-706-2

Perfect score: 1670

Sequence: 1 MTVTAINSONQKPKRLGL.....LFEFRISRGVYDPLTLK 322

Scoring table:

BLOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Ygapop 10.0 , Ygapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 1085931 segs, 780495707 residues

Total number of hits satisfying chosen parameters: 2171862

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPFO.spool/US10018706/runat_30062003_091107_23899/app_query.fasta.1.519
-DB=Published Applications NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOEXT=0 -UNITS=bite -START=1 -END=1 -MATRIX=blomsum62
-TRANS=human40.cdd -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=40 -MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MTNLEN=0
-MALLEN=2000000000 -USER=US10018706@cgn_1.1.118@runat_30062003_091107_23899
-NCPU=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSEPLCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:*

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PT1_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PTUTS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	Match length	DB	ID	Description
No.	Score				
1	219.5	13.1	1830121	9	US-10-329-960-1
2	209	12.5	2179	9	US-10-096-808-4
3	187	11.2	9025608	9	US-10-156-761-1
4	182.5	10.9	903	9	US-10-156-761-1629

5	175	10.5	609	9	US-10-156-761-4684	Sequence 4684, Ap
6	175	10.5	9025608	9	US-10-156-761-1	Sequence 1, Appl1
7	153	9.2	3353	9	US-09-927-827-41	Sequence 41, Appl
8	146.5	8.8	615	9	US-09-738-626-741	Sequence 2741, Ap
9	146.5	8.8	3309400	9	US-09-738-626-1	Sequence 1, Appl1
10	145	8.7	949	9	US-10-092-243A-8	Sequence 8, Appl1
11	136.5	8.2	1191	9	US-10-156-761-7229	Sequence 7229, Ap
12	135	8.1	1830121	9	US-10-329-960-1	Sequence 1, Appl1
13	130.5	7.8	15872	9	US-09-860-846-1	Sequence 1, Appl1
14	130.5	7.8	15872	9	US-09-860-846-1	Sequence 1, Appl1
15	130.5	7.8	15872	9	US-09-836-821-1	Sequence 1, Appl1
16	130.5	7.8	15872	10	US-09-836-821-1	Sequence 1, Appl1
17	128.5	7.7	15872	10	US-09-836-821-1	Sequence 1, Appl1
18	127	7.6	7277	9	US-09-468-147-164	Sequence 164, App
19	127	7.6	7277	9	US-09-468-147-165	Sequence 165, App
20	126	7.5	774	9	US-10-156-761-2407	Sequence 2407, Ap
21	125.5	7.5	942	10	US-09-770-445-354	Sequence 354, App
22	125	7.5	903	9	US-09-468-147-141	Sequence 141, App
23	122.5	7.3	2436	9	US-10-063-547-99	Sequence 99, Appl
24	122.5	7.3	2436	9	US-10-174-590-309	Sequence 309, Appl
25	122.5	7.3	2436	9	US-10-174-590-309	Sequence 309, Appl
26	122.5	7.3	2436	9	US-10-176-758-309	Sequence 309, Appl
27	122.5	7.3	2436	9	US-10-063-616-99	Sequence 99, Appl
28	122.5	7.3	2436	9	US-10-175-737-309	Sequence 309, Appl
29	122.5	7.3	2436	9	US-10-063-502-99	Sequence 99, Appl
30	122.5	7.3	2436	9	US-10-173-706-309	Sequence 309, Appl
31	122.5	7.3	2436	9	US-10-175-738-309	Sequence 309, Appl
32	122.5	7.3	2436	9	US-10-175-752-309	Sequence 309, Appl
33	122.5	7.3	2436	9	US-10-176-482-309	Sequence 309, Appl
34	122.5	7.3	2436	9	US-10-176-757-309	Sequence 309, Appl
35	122.5	7.3	2436	9	US-10-176-913-309	Sequence 309, Appl
36	122.5	7.3	2436	9	US-10-180-552-309	Sequence 309, Appl
37	122.5	7.3	2436	9	US-10-180-557-309	Sequence 309, Appl
38	122.5	7.3	2436	9	US-10-173-700-309	Sequence 309, Appl
39	122.5	7.3	2436	9	US-10-173-700-309	Sequence 309, Appl
40	122.5	7.3	2436	9	US-10-174-579-309	Sequence 309, Appl
41	122.5	7.3	2436	9	US-10-174-582-309	Sequence 309, Appl
42	122.5	7.3	2436	9	US-10-174-588-309	Sequence 309, Appl
43	122.5	7.3	2436	9	US-10-175-739-309	Sequence 309, Appl
44	122.5	7.3	2436	9	US-10-175-740-309	Sequence 309, Appl
45	122.5	7.3	2436	9	US-10-175-743-309	Sequence 309, Appl

ALIGNMENTS

RESULT 1
US-10-329-960-1
; Sequence 1, Application US/10329960
; Publication No. US20030099277A1
GENERAL INFORMATION:
APPLICANT: Fleischmann et al.
TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus influenzae Rd Genome, Fragm
FILE REFERENCE: PB186P1
CURRENT FILING DATE: 2003-01-02
CURRENT FILING DATE: 2003-01-02
PRIOR APPLICATION NUMBER: US 09/643,990
PRIOR APPLICATION NUMBER: US 08/487,429
PRIOR FILING DATE: 1995-06-07
PRIOR FILING DATE: 1995-06-07
PRIOR APPLICATION NUMBER: US 08/426,787
NUMBER OF SEQ ID NOS: 1
SOFTWARE: Patent version 3.1
SEQ ID NO 1
LENGTH: 1830121
TYPE: DNA
ORGANISM: Haemophilus influenzae
FEATURE:
NAME/KEY: misc feature
LOCATION: (4747)..(4747)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:

NAME/KEY: misc_feature
LOCATION: (9921)..(9921)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (10150)..(10150)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (29298)..(29298)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (36543)..(36543)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (36551)..(36551)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (36636)..(36636)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (40808)..(40810)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (44416)..(44416)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (44905)..(44905)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (44975)..(44975)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (45593)..(45593)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (45732)..(45732)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (47036)..(47036)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51334)..(51334)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51602)..(51602)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51786)..(51786)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51805)..(51805)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (55369)..(55369)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature

LOCATION: (65309)..(65309)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (65313)..(65313)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (80024)..(80024)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (100091)..(100091)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (102696)..(102696)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (105121)..(105121)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (107248)..(107248)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (117136)..(117136)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (119750)..(119750)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (119924)..(119924)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (120038)..(120038)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (121344)..(121344)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (122167)..(122167)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (122336)..(122336)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (131340)..(131340)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (131360)..(131360)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (139910)..(139910)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (140398)..(140398)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (142750)..(142750)

Qy	71	lenaaaglyglnvallysnglnsyaarprthvalserylalelagnargtryglsu	90
Db	749647	AATACCTTAAGAAGTAAACAAGGCCGATACATGTCTTATCGCTTAACCTGGGGGAT	749706
Qy	91	AsnTPraYglnlllelglyhisleasnaenleuasnserythriletrythrgly	110
Db	749707	GAATGAAAGAATGGCGCGCTTGATTAACCTATCCGAACCTTATTAATTAACTTTAGGG	749766
Qy	111	GlnTPleuthrLeuthrserglyasrleuylvalAlrglnkrgserlleeserSgly	130
Db	749767	CAAGTTTAAAGATTTCCTAATTTGTGCATPAAAAACGGTTACTACAAACGGTTTCT	749820
Qy	131	ValAsnThrAlaHisThrProserProValAlaValAlnSerSerArgProProValAln	150
Db	749821	CTAAAAACAACCTGCAGTCACAGCATCAACACCAACGCCAGTTAA	749865
Qy	151	GlnHisProAlaValAlnIlysrProThrPro	160
Db	749866	CCAGCAGTCACTTAT--ACTCCGGGGCGCAACCGTACTCAATAATAGATCTGAC	749918
Qy	161	ProValValValValValylsrProThrProThrProThrProProValVal	175
Db	749917	GGTACTATTATTGGCCCAATTAAATCAGAGGCTGGCAACCTGGTGTACTCTGGGCA	749976
Qy	176	GlnGlnProAlaProValAlaProProValAlThrGluAlaProPheAlaThrGlySerSer	195
Db	749977	ACTTCATCAACACAGGTTTCATCATCTCTGTAATAATAGCCGAATAGCACACCAATTAAATTC	750036
Qy	196	GlyValMetGlnPheArgTyrProValGlyAlaThrAsnProValValAlaArgAspGly	215
Db	750037	AATGTCGGTGGC--CCAAAT--GCCTCAAT--GTGGGTGGCAATGGCCG	750081
Qy	216	ThrAla-----ThrValAlaGlySerThrValThr-----SerArgGlyMetTrp	230

Db 1349 TTGAAATTC

	Pred. scores:				
	Pred. No.:	1.21e-12	Length:	2179	
	Score:	209.00	Matches:	63	
	Percent Similarity:	44.09%	Conservative:	49	
	Best local Similarity:	24.80%	Mismatches:	106	
	Query Match:	12.51%	Indels:	36	
	DB:	9	Gaps:	6	
US-10-018-706-2 (1-322) x US-10-098-808-4 (1-2179)					
Oy	73	TyrGlnValIysGlnGlyAspThrValSerIleLeuAlaArgTyrGlyLeuAsnTrp	92		
	::::: :::::::::::	::::: :::::::::::			
Db	1229	TACAAATATAGCAAGAAGCGAATACCATTCTTTCTTAATTCCTATATATTCCAGCATGGATATA	1288		
Oy	93	ArgGluIleGlyHisIleAsnAsnLysSerSerTyrThriIeYrThnGlyGIaTrp	112		
	::: :::::::::::	::: :::::::::::			
Db	1289	AAAACAATTTGCCACACACCAATTAATATGTCTGAGCCCAATATCATCTGACTATTGGCAAAGTA	1348		
Oy	113	LeuThrLeuTrpSerGly-----AspLeuIysValArgGluArgSerIleSerSer	129		
	::: :::::::::::	::: :::::::::::			
Db	1349	TTGAATAATTCGAATATATATTCCGATAGCATTGTGATTACCAACACAGACATTAAATGAA	1408		
Oy	130	GlyValAsnThrAlaHisIsthrProSer-ProValAlaValGlnSerSerArgProProVal	149		

Wed Jul 9 10:01:00 2003

us-10-018-706-2.rnpb

Page 4

```

Db      1409 TCAGAGGTGACCAAAATACATGCAATGAG--ACATGGAATCTTAATAACCAACAAAT 1465
QY      150 GlnGlnHisProAlaValGlnLysProThrProProValValValLysLysProThr 169
Db      1466 GAACCAATGAACCCGCTGTACACCAACATTCACCAATGCCAATCAATAA----- 1519
QY      170 ProThrProProValValGlnGlnProAlaProValAlaProProValThrGlnAlaPro 189
Db      1520 ---ACACCTCCACCAACCTCAATATATAGCTTGATTTGGCCA----- 1558
QY      190 PheAlaThrGlySerSerGlyValMetGlnPheArgTyrProValGlyAlaThrAsnPro 209
Db      1559 -----ACAAATGGAAAA 1570
QY      210 ValValArgArgPheGlyThrAlaThrValAlaGlySerThrValThrSerAsnGlyMet 229
Db      1571 ATTATTCAGAGATTTTCCAGTCTGATGGAGC-----ATAAAGGTATT 1615
QY      230 TrpPheSerGlyValArgAspGlyAspLeuIleAsnAlaSerAsnAlaGlyThrValIleGln 249
Db      1616 GATATTAGCGCTTCGTGTGACAGCTGTATTCAGACCTCTGACGAGTTGTATAT 1675
QY      250 AlaAspHisAsnMetAspGlyAlaSer-----IleValIleGlnHisThrAsnGlyPhe 267
Db      1676 GCCGAGACCGCTTACGTGATGTGTAATTATTAATTAACATATATGACAGTTAT 1735
QY      268 ValSerSerTyrIleHisIleLysAspAlaGlnValLysThrGlyAspThrValArgThr 287
Db      1736 TTAAGTCTATATGACATATGAAAGTATCTCTCAAGATCAGCAAGAGTTAAAGCG 1795
QY      288 GlyGlnArgGlyLeuAlaSerMetLysAsnGlnProSerGlyValAlaLeuPheGlnPheArg 307
Db      1796 GGTCAACAATCTCTAAATGGGAAGTCTGAAACAACAATCAAACTCCATTGTGA 1855
QY      308 IleSerArgAsnGlyValLysValAspProLeuThrValLeu 321
Db      1856 ATTCGTTATTAAGTCAATCAGTATGATCCAAATGAGATATTTA 1897

RESULT 3
US-10-156-761-1/c
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 349-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1
Alignment Scores:
Pred. No.: 3,23e-05 Length: 9025608
Score: 187.00 Matches: 88
```

```

Percent Similarity: 34.35% Conservative: 36
Best Local Similarity: 24.38% Mismatches: 117
Query Match: 11.20% Indels: 120
DB: 9 Gaps: 16

US-10-018-706-2 (1-322) x US-10-156-761-1 (1-9025608)
QY      27 ThrCysIleLeuAlaGlyCysAlaSerLysProThrTyrAsnSerThrSerGlySerGly 46
Db      2009653 ACAGCTACCTCGCGAGCATGAGAGAAATGACAC-----ATGCCCGCAAGAGGT 2009603
QY      47 SerHisArg-----ThrSerGlySerGly 54
Db      2009602 AAGCACCCCGCCAGAGTCCCGGTTTCAACCGCTCATGCGCGTCCGGAACCGGT 2009543
QY      55 GlyLeuAlaIle-----GlySerGlnValIleThrAspSer 66
Db      2009542 GCGCGCGGCTCGCGCTGCGCTGATGGGGGAGACCGGCGCCACCGCCGACCCGACG 2009483
QY      67 GlnGlyValProAsnArg-----GlyValValGlnGlyAspThrValSerTyrIleAlaGln 72
Db      2009482 GCGGCGCTCTCCGAAAGCTCTCCGACGCCCGCTGCGCGGAAGACAGGGGCGCGGAG 2009423
QY      73 -----TyrGlnValLysGlnGlyAspThrValSerTyrIleAlaGln 86
Db      2009422 AAGAGCGGACAGAGACGTAATGCGGTGGGGGGGTGCTCTCTCGAAGATCGCCAC 2009363
QY      87 ArgTyrGlyLeuAsnThrPArgGlnIleGlyHisIleAsnAsnLeuAsnSerSerTyrThr 106
Db      2009362 GAGCAGACGCTC----- 2009351
QY      107 IleTyrThrGlyGlnTyrLeuThrLeuTyrPheSerGlyAspLeuLysValArgGlyArgSer 126
Db      2009350 -----ACCGCGGCTGGAAGAGCTTACTCGACAAAC-----CGCTCC 2009312
QY      127 IleSerSerGlyValAsnThr-----AlaHis 135
Db      2009311 GCCATCGCGCGGACCCGACCTGATCCACCCCGGTGAAAGTCAAGATCGCGCCAG 2009252
QY      136 ThrProSerProValAlaValAlaGlnSerSerArgProProValGlnGlnHisProAlaVal 155
Db      2009251 TCCGCGTCAAGCGCGGCGCACCAAGTCGTC----- 2009222
QY      156 GlnLysProThrProProProValValValLysLysProThrProThrProProValVal 175
Db      2009221 ---ACGGGAGAGAGCCGCGCACGCGGCTGAAGTCCGCGACG-----GCG 2009180
QY      176 GlnGlnProAlaProValAlaProProValThrGlnAlaProPheAlaThrGlySerSer 195
Db      2009179 AAGACCCCGCGGTGGAAGACGACACCGGCCCGCGGCGGACACCAACCGCGGCC 2009120
QY      196 GlyValMetGlnPheAsnGlyTyrProValGlyAlaThrAsnProValValArgArgPheGly 215
Db      2009119 GGC-----TACCCCTCCGCTGACGCGC----- 2009096
QY      216 ThrAlaThrValAlaGlySerThrValThrSerAsnGlyMetTyrPheSerGly----- 233
Db      2009095 ---GCCACCATCGGTACCGCTTACAGACCGCGGACAGCATGTGTCTCCAGGGGTACCAC 2009039
QY      234 -----ArgAspGlyAspLeuIleAsnAlaSerAsnAlaGlyThr 246
Db      2009038 ACCGCGCTGCACTTCGTGTCGCCACCGGCAACCAATCAAGCGCGTCCGCGCGGACAC 2008979
QY      247 ValIleGlnAlaAsp---HisAsnMetAspGlyAlaSerIleValIleGlnHisThrAsn 265
Db      2008978 GTCGTCGCGCGCGCTGGGCGGCGCGGACGCAACGAGTCTGTCGCGCGGACGCGCAT 2008919
QY      266 GlyPheValSerSerTyrIleHisIleLysAspAlaGlnValLysThrGlyAspThrVal 285
Db      2008918 GCGCAGTACTCGCAGTACGACCAATGTCACAGCTCCCGTCCACAGGGGCGAGAGCGTG 2008859
QY      286 ArgThrGlyGlnArg-----IleAlaSerMetLysAsnGlnProSerGlyValAlaLeu 303
```

Db 2008858 GCCGAGGCGCGCAGCTGGGCTCTGTGTGAGACCGGCAATGTCAACGGAGCCGACCTGT 2008799
Qy 304 ---PheGluPheArgIleSerArgAsn---GlyValTyrValAspProLeuThrValLeu 321
Db 2008798 CACTTCGAGATCCGTAGACACCCCTCTTACGCTCGAGCGTGAACCGGCTGCTTACCTG 2008739
Qy 322 Lys 322
Db 2008738 CGC 2008736
RESULT 4
US-10-156-761-1629
; Sequence 1629, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMIURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1629
; LENGTH: 903
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(903)
US-10-156-761-1629
Alignment Scores:
Pred. No.: 2,52e-10 Length: 903
Score: 182.50 Matches: 83
Percent Similarity: 34.50% Conservative: 35
Best Local Similarity: 24.27% Mismatches: 107
Query Match: 10.93% Indels: 117
DB: 9 Gaps: 15
US-10-018-706-2 (1-322) x US-10-156-761-1629 (1-903)
Qy 46 GlySerHisArg-----ThrSerGlySer 53
Db 13 GGTAAACACACCGCCGCGCCAGTCCCGCGTTTACCGCTGATGCGCGCGGAAC 72
Qy 54 GlyGlyLeuAlaIle-----GlySerGlnValIleThrAsp 65
Db 73 GGTGGCGCGCGCTCGCGCTCCGCTGATGGGGGCGAGCGCGCCACGCCGCCACCCG 132
Qy 66 SerGlnGlyValProAsnArg----- 72
Db 133 ACGGCGCGCTCTCCGGAAGGTCTCCGACGCCCGCTTCCGCGGAAGCAGGGGCGCGC 192
Qy 73 -----TyrGlnValysGlnGlyAspThrValSerIleIleAla 85
Db 193 GAGAAAGCGCGCAGCAGACGTATGCGGTGCGGGGCTGACTCCCTTCGAAGATCGCC 252
Qy 86 GlnArgTyrGlyLeuAsnTyrArgIleGlyHisIleAsnLeuAsnSerSerTyr 105
Db 253 GACGACGACGAGCTC----- 267
Qy 106 ThrIleTyrThrGlyGlnTyrLeuThrLeuTyrSerGlyAspLeuLysValArgGlnArg 125
Db 268 -----ACCGGCGCGCTGGAGAGAGCTTACTCTGACCAAC-----CGC 303

Qy 126 SerIleSerSerGlyValAsnThr-----Ala 134
Db 304 TCCGCCATCGCGCGCGACCGGACCGCTGATCCACCCCGGTCTGAAGCTGACATGCGGCC 363
Qy 135 HisThrProSerProValAlaValGlnSerSerArgProProValGlnHisProAla 154
Db 364 AAGTCCGCTCGAGCGCGCCGACGACGAGTCTGCTC----- 396
Qy 155 ValGlnIleProThrProProValValValValValValValValValValValValVal 174
Db 397 -----ACGGCGAGAGAGCGCGCGCCGCGGAGTGAAGTCCGCCACG----- 435
Qy 175 ValGlnGlnProAlaProValAlaProProValThrGlnAlaProPheAlaThrGlySer 194
Db 436 GCGAAGACCCCGCGGTGAGAGACACACACCGCCCGCGCGGACACACACACCGCGC 495
Qy 195 SerGlyValMetGlnPheArgTyrProValGlyValThrAsnProValValArgAspPhe 214
Db 496 GCCGCG-----TACACCTGCGCGTGAACGCG----- 522
Qy 215 GlyThrAlaThrValAlaGlySerThrValThrSerAsnGlyMetTyrPheSerGly--- 233
Db 523 -----GCCACATCGGTACCGCTTACAGACCGCGGACGATGTGTCCAGCGGTAC 576
Qy 234 -----ArgAspGlyAspLeuIleAsnAlaSerAsnAlaGly 245
Db 577 CACACCGCGCTGACTTGTGTGTCGCCACCGGACACCATCAAGCCGTGCGCGCGCGC 636
Qy 246 ThrValIleGlnAlaAsp---HisAsnMetAspGlyAlaSerIleValIleGlnHisThr 264
Db 637 ACGCTGCTCGCGCGCGGTGCGCGCGCGCGGTGACGACGAGTGTGTGCGCGCACGCC 696
Qy 265 AsnGlyPheValSerSerTyrIleHisIleValAspAlaGlnValIleThrGlyAspThr 284
Db 697 GATGCCAGTACTGCGAGTACGCGGACCATGTCCAGCTTCCGTCCACGCGGACGACG 756
Qy 285 ValArgThrGlyLeuArg-----IleAlaSerMetLysAsnGlnProSerGlyValAla 302
Db 757 GTGGCCGAGGCGCGGACGCTGCGCTTGTGTGCGACCGGCAATGTCAACGGAGCCGAC 816
Qy 303 Leu---PheGluPheArgIleSerArgAsn---GlyValTyrValAspProLeuThrVal 320
Db 817 CTGCACTTCGAGATCCGTAGACACCCCTCTTACGAGTCTCGAGCGTGAACCGGCTGCTTAC 876
Qy 321 LeuLys 322
Db 877 CTGCGC 882
RESULT 5
US-10-156-761-4684
; Sequence 4684, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMIURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 4684
; LENGTH: 609
; TYPE: DNA

Wed Jul 9 10:01:00 2003

us-10-018-706-2.rnpb

Page 6

```

; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(609)
US-10-156-761-4684

Alignment Scores:
Pred. No.: 9,31e-10 Length: 609
Score: 175.00 Matches: 66
Percent Similarity: 40.69% Conservative: 17
Best Local Similarity: 32.35% Mismatches: 19
Query Match: 10.48% Indels: 53
DB: Gaps: 10

US-10-018-706-2 (1-322) x US-10-156-761-4684 (1-609)

QY 137 ProSerProValAlaValAlaGlnSerSerArgProProValGlnGlnHisProAlaValAlaGln 156
DB CCTCGGCGCTGCTGGCGCGCGGCTCCCTCCGCG----- 118

QY 157 LysProThrProProValAlaValAlaValLysLysProThrProProProValAlaGln 176
DB 119 ---CGACACCG-----GGCTCCAGCGCAGTGGCGCGGCGCGCGTGC 157

QY 177 GlnProAlaProValAlaProProValThrGlnAlaProPheAlaThrGlySerSergly 196
DB 158 AGACCCAGCGCGCGCGCGCA---AGAGCGCGCGCTCCCTCGATGATGACCCGG 211

QY 197 ValMetGlnPheArgTyrProValGlyAlaThrAsnProValAlaArgArgPheGlyThr 216
DB 212 TCAGAGATAGCGGCTCTCCGCGCGCTTCAACAGCGCGC----- 252

QY 217 AlaThrValAlaGlySerThrValThrSerAsnGlyMetTrp-----PheSergly 233
DB 253 -----GGCATGTGGCGGCCCAACAGCATCCGGT 279

QY 234 ArgAsp-----GlyAspLeuIleAsnAlaSerAsnAlaGlyThrValIle 248
DB 280 CAGACCTTCCCGCGCCAGCGGTAACGAGTCTCCCGCGCGCGCGCGCGCGCGTGC 339

QY 249 GlnAlaAspHisAsn-----MetAspGlyAlaSerIleValIleGln 262
DB 340 AAGCGCGCGCGCAACCGCGCGCGTACGCGCGCGCGCGCGCGCGCGCGCGTCAAG 399

QY 263 HisThrAsnGlyPheValSerSerTyrIleHisIleLysAspAlaGlnValLysThrGly 282
DB 400 CACGCGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 459

QY 283 AspThrValArgThrGlyGlnArgIleAlaSerMetLysAsn-----GlnProSergly 300
DB 460 CAGGTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 519

QY 301 AlaAlaLeu---PheGlnPheArgIleSerArgAsn---GlyValTyrValAspProLeu 318
DB 520 CCGCAGCTGCACTTGCAGATCGTACGACCCCGAATACGAGTCCGCGCGCGCGCGT 579

QY 319 ThrValLeuLys 322
DB 580 GCCTTCTCCGCG 591

RESULT 6
US-10-156-761-1
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMIKA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
```

```

; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Alignment Scores:
Pred. No.: 0.000633 Length: 9025608
Score: 175.00 Matches: 66
Percent Similarity: 40.69% Conservative: 17
Best Local Similarity: 32.35% Mismatches: 19
Query Match: 10.48% Indels: 53
DB: Gaps: 10

US-10-018-706-2 (1-322) x US-10-156-761-1 (1-9025608)

QY 137 ProSerProValAlaValAlaGlnSerSerArgProProValGlnGlnHisProAlaValAlaGln 156
DB 5728773 CCTCGGCGCTGCTGGCGCGCGGCTCCCTCCGCG----- 5728808

QY 157 LysProThrProProValAlaValAlaValLysLysProThrProProProValAlaGln 176
DB 5728809 ---CGACACCG-----GGCTCCAGCGCAGTGGCGCGGCGCGCGTGC 5728847

QY 177 GlnProAlaProValAlaProProValThrGlnAlaProPheAlaThrGlySerSergly 196
DB 5728848 AGACCCAGCGCGCGCGCGCA---AGAGCGCGCGCTCCCTCGATGATGACCCGG 5728901

QY 197 ValMetGlnPheArgTyrProValGlyAlaThrAsnProValAlaArgArgPheGlyThr 216
DB 5728902 TCAGAGATAGCGGCTCTCCGCGCGCTTCAACAGCGCGC----- 5728942

QY 217 AlaThrValAlaGlySerThrValThrSerAsnGlyMetTrp-----PheSergly 233
DB 5728943 -----GGCATGTGGCGGCCCAACAGCATCCGGT 5728969

QY 234 ArgAsp-----GlyAspLeuIleAsnAlaSerAsnAlaGlyThrValIle 248
DB 5728970 CAGACCTTCCCGCGCCAGCGGTAACGAGTCTCCCGCGCGCGCGCGCGCGTGC 5729029

QY 249 GlnAlaAspHisAsn-----MetAspGlyAlaSerIleValIleGln 262
DB 5729030 AAGCGCGCGCGCAACCGCGCGCGTACGCGCGCGCGCGCGCGCGCGCGG 5729089

QY 263 HisThrAsnGlyPheValSerSerTyrIleHisIleLysAspAlaGlnValLysThrGly 282
DB 5729090 CACGCGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 5729149

QY 283 AspThrValArgThrGlyGlnArgIleAlaSerMetLysAsn-----GlnProSergly 300
DB 5729150 CAGGTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 5729209

QY 301 AlaAlaLeu---PheGlnPheArgIleSerArgAsn---GlyValTyrValAspProLeu 318
DB 5729210 CCGCAGCTGCACTTGCAGATCGTACGACCCCGAATACGAGTCCGCGCGCGCGT 5729269

QY 319 ThrValLeuLys 322
DB 5729270 GCCTTCTCCGCG 5729281

RESULT 7
```

```

US-09-927-827-41
; Sequence 41, Application US/09927827
; Publication No. US20030036176A1
; GENERAL INFORMATION:
; APPLICANT: Bower, Stanley G.
; APPLICANT: Rameeler, Thomas M.
; TITLE OF INVENTION: Directed Genetic Engineering of Xanthomonas campestris
; FILE REFERENCE: 38-10(15824)B
; CURRENT APPLICATION NUMBER: US/09/927,827
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: US 60/279,493
; PRIOR FILING DATE: 2001-03-28
; NUMBER OF SEQ ID NOS: 69
; SEQ ID NO 41
; LENGTH: 3353
; TYPE: DNA
; ORGANISM: Xanthomonas campestris
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1001)..(2353)
US-09-927-827-41

Alignment Scores:
Pred. No.: 2,366-06 Length: 3353
Score: 153.00 Matches: 61
Percent Similarity: 42.47% Conservative: 32
Best Local Similarity: 27.85% Mismatches: 85
Query Match: 9.16% Indels: 42
DB: 9 Gaps: 11

US-10-018-706-2 (1-322) x US-09-927-827-41 (1-3353)
Qy 111 GlnTPrLeuThrLeuTPrSerGlyAspLeuLysValAlaGlnArgSerIleSerSerGly 130
Db 46 CGCTGAGACAGTGTGCTGCGCAATCTGCTGACGCGCGGCAAGCGCG----- 93
Qy 131 ValAenThAlaHisThrProSerProValAlaValGlnSerSerArgProPro--Val 149
Db 94 -----ACGCCGAGCGCGCGCGCGCGCAAGCGTGGCGCGCGCAAGCCG 138
Qy 150 GlnGlnHisProAlaValGlnLysProThrProProValValValLysLysProThr 169
Db 139 CGCGCGAGGCCAAGCGCGCGCAAAACCGATCGC-CGC-----GACGCCCGCGC 185
Qy 170 ProThrProProValValGlnGlnProAlaProValAlaProProValThrGlnAlaPro 189
Db 186 AAGACTCGCGCCCAAGGTGCGGCAACGACCG-----GGGCCA 224
Qy 190 PheAlaThrGlySerSerGlyValMetGlnPheArgTYrProValGlnAlaThrAsnPro 209
Db 225 AAGTCGTGGTGGCTTGGT-----TGGCGGTGTCTGGGCAAT----- 260
Qy 210 ValValArgArgPheGlyThrAlaThrValAlaGlySerThrValThrSerAsnGlyMet 229
Db 261 CTGCTGGCGCGCTTCAACGCCACGCTGCCGATGGCCAT-----ACCGCAAGGGCGTG 314
Qy 230 TrpPheSerGlyArgAspGlyAspLeuIleAsnAlaSerAsnAlaGlyThrValIleGln 249
Db 315 CTGATCGGGCGGCCCAAGGGGACAGACGTCACCGCGGTGGCGATGGCAACGGTGGTTT 374
Qy 250 AlaAspHisAsnMetAsp--GlyAlaSerIleValIleGlnHisThrAsnGlyPheVal 268
Db 375 TCCGACTGGATGACCGGCTACCGCATGATCTCGATCTGTGATATCAGCGCATGGCTCATAG 434
Qy 269 SerSerTYrIleHis-----IleLysAspAlaGlnValLysThrGlyAspThr 284
Db 435 ACCCTGTACGGCCACACACACACTTGTCTGGCGGATGCC-----GGCGCCAGC 482
Qy 285 ValArgThrGlyGlnArgIleAlaSerMet--LysAsnGlnProSerSerGlyAlaAlaLeu 303
Db 483 ATCAAAACGCGGAGACCGGTGGCAAAAGTCGGAGTTCGGGCGGGGACAGGTGTGCGGGCG 542
Qy 304 PheGlnPheArgIleSerArgAsnGlyValTYrValAspProLeuThrValLeuLys 322

```

RESULT 8

US-09-738-626-2741
; Sequence 2741, Application US/09738626
; Publication No. US20020197605A1
GENERAL INFORMATION:
APPLICANT: MATSUYAMA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: OCHIAI, KEIKO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: YATEISHI, NAKO
APPLICANT: SENOH, AKIHITO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738,626
CURRENT FILING DATE: 2000-12-18
PRIORITY APPLICATION NUMBER: JP 99/377484
PRIORITY FILING DATE: 1999-12-16
PRIORITY APPLICATION NUMBER: JP 00/159162
PRIORITY FILING DATE: 2000-04-07
PRIORITY APPLICATION NUMBER: JP 00/280988
PRIORITY FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 2741
LENGTH: 615
TYPE: DNA
ORGANISM: Corynebacterium glutamicum
US-09-738-626-2741

Alignment Scores:
Pred. No.:
Score: 1.1e-06 Length: 615
Percent Similarity: 146.50 Matches: 64
Best local Similarity: 44.44% Conservative: 20
Query Match: 33.86% Mismatches: 67
DB: 8.77% Indels: 39
Gaps: 10

US-10-018-706-2 (1-322) x US-09-738-626-2741 (1-615)

Oy	137	ProSerProValAlaValAlaGlnSerSerArgrProValAlaGlnGlnIlnHIsProAlaValAlaGln	156
Db	107	CCGCACCAACCGCTAACCG-----CCACCATGGACCGCAGCATCCG-----CGC	145
Oy	157	LysProThr-----ProProValAlaValAlaValLysLysProThrProThrPro	172
Db	146	AACCGACTACGCGAGCGCTTCTCTCGGCCGTG-----CGACACCGCTCCAG	139
Oy	173	ProValAlaGln--GlnProAlaProValAlaProProValThnrglAlaProPhoAla	190
Db	194	AAGCTCACGAGAGTCGACACACCGCGCTC-GCACCGCGCGCACCGTGAGCGCGCCAGCA	252
Oy	192	ThnrglySerSerGlyValMetGlnHeargtryrProValGlyAlaThraenProValAl	211
Db	253	AACGGGACC-----TTACCTCAGGATTCCGA	279
Oy	212	ArgArgPheGlyThrAlaThrValAlaGlySerThrValThrSerAnnglyMetTrpPhe	231
Db	280	CCAGCTTGGGGAAC-----TTCCACAACGGCATCGACATC	315
Oy	232	SerGlyArgAspGlyAspleulleasnAlaserAsnAlaglyThrVallleglnAlaasp	251
Db	316	GCAAACTCATATGGGCAACCCCATTCACCCGCTCATATGGCGGCGCATGTCATCAGCTCGC	375
Oy	252	HisaMetserasp---GlyAlaserlleVallleglnHIsThrAnnglyPheValSerSer	270
Db	376	CCAGCATCCGGCTATAGAACTGGATGGATCCGATCCAGCAAGCGATCATCTCCATC	435

QY 271 TTTTLEHISLELYASAPALAGINVALYSTRGJYASPTHYVALARGTHRGJYGINARG 290
DB 436 TACGACACATGGAGATTCCTTACGTCCTCGCGCGACGCGTCGACGACGACGAGAA 495
QY 291 TLEALASERMELYASANGINPRO--SergiyAlaIaleuphegluphearglleser 309
DB 496 ATCGACGAGATGGACAGCCAGGATTCCTCCACCGGCTCCACCTTCGATTCGATTCAC 555
QY 310 ARGASNGLYVALTYR--ValAspPro 317
DB 556 CCAGACGGCGTCACCCCGACGTCGACCCA 582

RESULT 9
US-09-738-626-1
Sequence 1, Application US/09738626
Publication No. US20020197605A1
GENERAL INFORMATION:
APPLICANT: NAKAGAWA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, MOKO
APPLICANT: SENOH, AKIHIRO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738, 626
CURRENT FILING DATE: 2000-12-16
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 1
LENGTH: 3309400
TYPE: DNA
ORGANISM: Corynebacterium glutamicum
US-09-738-626-1

Alignment Scores:
Pred. No.: 0.182 Length: 3309400
Score: 146.50 Matches: 64
Percent Similarity: 44.44% Conservative: 20
Best Local Similarity: 33.86% Mismatches: 29
Query Match: 8.77% Indels: 39
DB: 9 Gaps: 10

US-10-018-706-2 (1-322) x US-09-738-626-1 (1-3309400)

QY 137 ProSerProValAlaValGlnSerArgProProValGlnGlnHisProAlaValGln 156
DB 2649656 CCGCACCGCTACCG-----CCACCATGGCACACGATTCG-----CGC 2649694
QY 157 LysProThr-----ProProValAlaValAlaLysProThrProThrPro 172
DB 2649695 AAACCCGATCGACGAGCTTCTCCCGCGCGTGG-----CCGACACCGTCGCGAG 2649742
QY 173 ProValValGln--GlnProAlaProValAlaProProValThrGlnAlaProPheAla 191
DB 2649743 AACCTGACGAGATCGACACCGCGCTTC--GCACCAACCGCCCGAGGCGCCGACGCA 2649801
QY 192 ThrGlySerSerGlyValMetGlnPheArgTyrProValGlyAlaThrAsnProValAl 211
DB 2649802 AACGGCAC-----TTCACTCGAGATTCGGA 2649828
QY 212 ArgArgPheGlyThrAlaThrValAlaGlySerThrValThrSerAsnGlyMetTrpPhe 231

DB 2649829 CCACGTTGGGAGACC-----TTCCACACCGCATCGACATC 2649864
QY 232 SerGlyArgAspGlyAspLeuIleAsnAlaSerAsnAlaGlyThrValIleGlnAlaAsp 251
DB 2649865 GCAAATCAATCGGACCCCAATTCACCTGATGCGGACGATGATGATGATGATGATGATG 2649924
QY 252 HisAsnMetAsp--GlyAlaSerIleValIleGlnHisThrAsnGlyPheValSerSer 270
DB 2649925 CCAGATCCGCTATGACAGTGCATTCGACGACGACGACGACGATTCATTCATC 2649984
QY 271 TTTTLEHISLELYASAPALAGINVALYSTRGJYASPTHYVALARGTHRGJYGINARG 290
DB 2649985 TACGACACATGGAGATTCCTTACGTCCTCGCGCGACGCGTCGACGACGACGAGAA 2650044
QY 291 TLEALASERMELYASANGINPRO--SergiyAlaIaleuphegluphearglleser 309
DB 2650045 ATCGACGAGATGGACAGCCAGGATTCCTCCACCGGCTCCACCTTCGATTCGATTCAC 2650104
QY 310 ARGASNGLYVALTYR--ValAspPro 317
DB 2650105 CCAGACGGCGTCACCCCGACGTCGACCCA 2650131

RESULT 10
US-10-092-243A-8/c
Sequence 8, Application US/10092243A
Publication No. US20020197625A1
GENERAL INFORMATION:
APPLICANT: Hillman, Jeffrey D.
TITLE OF INVENTION: Microbial polynucleotides Expressed During Infection of
FILE REFERENCE: MEH800-5051
CURRENT APPLICATION NUMBER: US/10/092, 243A
CURRENT FILING DATE: 2002-05-21
PRIOR APPLICATION NUMBER: 60/147,551
PRIOR FILING DATE: 1999-08-06
PRIOR APPLICATION NUMBER: US00/21340
PRIOR FILING DATE: 2000-08-04
NUMBER OF SEQ ID NOS: 20
SOFTWARE: PatentIn ver. 2.1
SEQ ID NO 8
LENGTH: 949
TYPE: DNA
ORGANISM: Actinobacillus actinomycetemcomitans
FEATURE:
NAME/KEY: misc feature
LOCATION: (538)
OTHER INFORMATION: N stands for any nucleotide.
US-10-092-243A-8

Alignment Scores:
Pred. No.: 2.93e-06 Length: 949
Score: 145.00 Matches: 57
Percent Similarity: 37.04% Conservative: 23
Best Local Similarity: 26.39% Mismatches: 62
Query Match: 8.68% Indels: 74
DB: 9 Gaps: 9

US-10-018-706-2 (1-322) x US-10-092-243A-8 (1-949)

QY 73 TTTGlnValLysGlnGlyAspThrValSerIleAlaGlnArgTyrGlyLeuAsnTrp 92
DB 537 TTTTACCGTACGCAAGGAGGACACCAATGTTCTTATCGCTTCATTCGTTTGGATGTG 478
QY 93 ArgGlnIleGlyHisIleAsnAsnLeuAsnSerSerTyrThrIleTyrThrGlyGlnTrp 112
DB 477 AAAGATTTGGCGCGCTTAAATTAATATGTCGCAAACTTACCTTAAGGTCGACCAACC 418
QY 113 LeuThrLeuTrpSerGlyAspLeuIleValArgGlnArgSerIleSerSerGlyValAsn 132
DB 417 TTA-----AAAGTACGCGCGCGC----- 400
QY 133 ThrAlaHisThrProSerProValAlaValGlnSerSerArgProProValGlnGlnHis 152


```

Db      399  -----ACACGGTTACGGTGCACAAAGCTGACCCCAATGCA--- 358
Qy      153  ProAlaValGlnLysProThrProProValValValLysLysProThrProThrPro 172
Db      357  -----CCACCGGTCCACCCAA 343
Qy      173  ProValValGlnGlnProAlaProProValAlaProProValAlaProPheAlaThr 192
Db      342  CCGGTTACGCAAGGTGGTGAACCGAC---GTACTTACACCCCGGCGCAAC 292
Qy      193  GlySerSerGlyValMetGlnPheArgTyrProValGlyAlaThrAsnProValAlaArg 212
Db      291  GGCAACA-----CAATATGTTCCGACGCGACGATTAACCGCCGATTAAGCC 244
Qy      213  ArgPheGlyThrAla-----ThrValAlaGlySerThr--- 223
Db      243  GCGGTAGGTACTGTCGACCGCGCAACCAATCAGCCGATTAACCGGCGACCGCACCG 184
Qy      224  ValThrSerArgLysMetTrp-----PheSer 232
Db      183  GCAGTATCCAAACGTGGACGTGCGCTGCGCCGACCAAGGCAATGTGATTCAAGTTTCTCT 124
Qy      233  GlyArgAspGly-----AspLeuIleAsn 240
Db      123  AACGCCGACGCGGTAAACAAAGCATGCATCAGTGTCTTCGACGACAAAGCATCAT 64
Qy      241  AlAserAsnAlaGlyThrValIleGlnAlaAspHisAsnMetAspGly 256
Db      63  GCCCGCGCGCGAGTGTGATTACCGCGGTAAACGCTTACCGCGGT 16

```

RESULT 11

```

US-10-156-761-7229
; Sequence 7229, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 7229
; LENGTH: 1191
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1191)
US-10-156-761-7229

```

Alignment Scores:

```

Pred. No.: 3,31e-05 Length: 1191
Score: 136.50 Matches: 82
Percent Similarity: 36.70% Conservative: 27
Best Local Similarity: 27.61% Mismatches: 126
Query Match: 8.17% Indels: 63
DB: 9 Gaps: 15

```

US-10-018-706-2 (1-322) x US-10-156-761-7229 (1-1191)

```

Qy      43  SerGlySerGlySerHisArgThrSerGlySerGlyGlyLeuAlaIleGlySer----- 60
      |||||
      :|:|:|
      :|:|

```

```

Db      421  TCCGCGACGCGACACACACCGCGGACCGAGACCAAGCGGACGTCCTGATGTGAG 480
Qy      61  ---GlnValIleThrAspSerGln---GlyValProAsnArgTyrGlnValLysGlnGly 78
Db      481  CCACAGACGACACCGCAGTCGACAGCGGCGGACCGCCGAGATGTACCGCTTCGACGCG 540
Qy      79  AspThrValSerLysIleAlaGlnArgTyrGlyLeuAsnTrpArgGlnIleGlyHisIle 98
Db      541  GACACGCTCTCCGGATCCG---GACTCCCGGACAGTC 576
Qy      99  AsnAsn-LeuAsnSerSerTyrThrIleTyrThrGlyGlnTrpLeuThrLeuTrpSerG 118
Db      577  CGGCGCGGTGGAGCAGCTCTACGCGCCACCGCG---GGACC 618
Qy      118  YAspLeuLysValArgLysArgSerIle---SerSerGlyValAsnThrAlaHisTrp 137
Db      619  ATCGCGCCGACCGCGACCTGATCTTCGCGGACGCGGCTGATCTGCGGCGCAAGCGC 678
Qy      137  oSerProValAlaValGlnSerSerArgProProValGlnGlnHisProAlaValGlnLys 157
Db      679  ACGACCCGACCGACGACGAGCGCTCCAGCTCCAGCCCAAGTCG--- 721
Qy      157  sProThrProProValValValValLysLysProThrProThrProProValValGlnG 177
Db      722  -----CGACGCTCTCTCAAGTCTCGACGCGAAGAGCG 756
Qy      177  nProAlaProValAlaProProValAlaThrGlnAlaProPheAlaThrGlySerSerGly 197
Db      757  TCTCTCG---ACAGACCG---AGAGCGGACGACGAGCCACTCTCTGTCGCGC--- 804
Qy      197  lMetGlnPheArgTyrProValGlyAlaThrAsnProValValArgArgPheGlyThrAl 217
Db      805  -----CCGCTGAACGCC-----GGACCGG 824
Qy      217  aThr-----ValAlaGlySerThrValThrSer-----AsnGlyMetTrpPh 231
Db      825  GACCAATATCCATCGCGCGGCTTCTCGCTCGATCGAAGGGCTATCACACCGCGCTGACTT 884
Qy      231  eSerGlyArgAspGlyAspLeuIleAsnAlaSerAsnAlaGlyThrValIleGlnAlaAs 251
Db      885  CCCCCTGCCACCGGCGACCTCCGCGAAGCGGTCGCGCGGCGGACGTCGTCACCGCGG 944
Qy      251  pHisAsnMetAsp---GlyAlaSerIleValIleGlnHisThrAsnPheValSerSe 270
Db      945  CTGGGCGCGCTCTCTCGCTTACCAAGTGTCTCTCGGACCGCGACCGCGCTTACCGCA 1004
Qy      270  rTyrIleHisIleLysAspAlaGlnValLysThrGlyAspThrValArgThrGlyGlnAr 290
Db      1005  GTACGCCCATCTGTGCGCATCTCCGTGAAGCGCGGCGACAGCGCGCACCGCG 1064
Qy      290  gIle-----AlaSerMetLysAsnGlnProSerGlyAlaAlaLeuPheGlyLysPheAr 307
Db      1065  TATCGCGCGCTCGGCGGTCCACGCGGCAACAGACGCGCGCGCATCTGCACTTCGAGTGG 1124
Qy      307  gIleSerArgAsn---GlyValTyrValAspProLeuThrValLeuLys 322
Db      1125  GACGCGCGCGCGCTTCCGCGAGTGCATCGATCCGCTCATCTGAGG 1173

```

RESULT 12

```

US-10-329-960-1/c
; Sequence 1, Application US/10329960
; Publication No. US20030099277A1
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus influenzae Rd Genome, Fragm
; FILE REFERENCE: P186P1
; CURRENT FILING DATE: 2003-01-02
; PRIOR APPLICATION NUMBER: US 09/643,990
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: US 08/487,429
; PRIOR FILING DATE: 1995-06-07

```

PRIOR APPLICATION NUMBER: US 08/426,787
PRIOR FILING DATE: 1995-04-21
NUMBER OF SEQ ID NOS: 1
SOFTWARE: PatentIn version 3.1
SEQ ID NO 1
LENGTH: 1830121
TYPE: DNA
ORGANISM: Haemophilus influenzae
FEATURE:
NAME/KEY: misc_feature
LOCATION: (4747)..(4747)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (9921)..(9921)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (10150)..(10150)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (29258)..(29298)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (36543)..(36543)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (36551)..(36551)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (36636)..(36636)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (40808)..(40810)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (44416)..(44416)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (44905)..(44905)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (44975)..(44975)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (45393)..(45393)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (45732)..(45732)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (47036)..(47036)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51334)..(51334)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51602)..(51602)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:

NAME/KEY: misc_feature
LOCATION: (51786)..(51786)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51805)..(51805)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (55369)..(55369)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (65309)..(65309)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (65313)..(65313)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (80024)..(80024)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (100091)..(100091)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (102696)..(102696)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (105121)..(105121)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (107248)..(107248)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (117136)..(117136)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (119750)..(119750)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (119924)..(119924)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (120038)..(120038)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (121344)..(121344)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (122167)..(122167)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (122336)..(122336)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (131340)..(131340)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature

```
LOCATION: (131360)..(131360)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (139910)..(139910)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (140398)..(140398)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (142750)..(142750)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (145058)..(145058)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (145171)..(145171)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (145942)..(145942)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (147197)..(147197)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (150841)..(150841)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (152500)..(152500)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (152530)..(152530)

Alignment Scores:
Pred. No.: 1.37 Length: 1830121
Score: 135.00 Matches: 63
Percent Similarity: 39.29% Conservative: 36
Best Local Similarity: 25.00% Mismatches: 111
Query Match: 8.08% Indels: 42
DB: Gaps: 10

US-10-018-706-2 (1-322) x US-10-329-960-1 (1-1830121)
QY 74 GlnValysGlnGlyAspThrValSerIleAlaGlnArgTyrGlyLeuAsnTrpArg 93
Db 820115 CAAAAAACAACAACAAGCATTCGAAAAACACAGCAGCATCAA----- 820068
QY 94 GluIleGlyHisIleAsnAsnLeuAsnSerSerTyrTrpIleTyrThrGlyGlnTrpLeu 113
Db 820067 -----TCTACGCTGGAATGAATCAATTAATAATTAGCCCTGATCAAGATTAATTGAAT 820014
QY 114 ThrLeuTrpSerGlyAspLeuIleValArgGluArgSerIleSerSerGlyValAsnThr 133
Db 820013 ACA-----CTAAAGCAACAGAACAGACACTTCGCAAGAAATTCACGA 819969
QY 134 AlaHisThrProSerProValAlaValGlnSerSerArgProProVal---GlnGlnHis 152
Db 819968 GCTGAACAAGCAGCGCGCAACAAGAAACGTAAGAGAGAGCAGCTTCCTCAAGCCAA 819909
QY 153 ProIleValGlnIleProThrProProValValValValValValValValValValVal 172
Db 819908 AAGGCTGAAGAAACAACA----- 819888
QY 173 ProValValGlnGlnProIleProValAlaProValAlaProValThrGluAlaProPheAlaThr 192
```

```
Db 819887 -----TCAAAACCTTATCAACCAACTGTGCAGAACGCCAATTACTTAAT 819843
QY 193 GlySerSerGlyVal-----MetGlnPheArgTyrProValGlyAlaThrAsn 208
Db 819842 AGTACAAAGCGGTTTAAAGGCGCGCAAAAAACAATATCTTACCAGTTCTGCTTCA--- 819786
QY 209 ProValValArgArgPheGlyThrAlaThrValAlaGlySerThrValThrSerAsnGly 228
Db 819785 ---ATTTCGATCTTTTGGT-----TCTATCCAAACAGCGCAAGTACGTGGAAAGGT 819735
QY 229 MetTrpPheSerGlyArgAspGlyAspLeuIleAsnAlaSerAsnAlaGlyThrValIle 248
Db 819734 ATGGTAATTCGCGCATCAGCAGCAGCGCTGTAAACCAATTCGCTGCAGCGCATTT 819675
QY 249 GlnAlaAspHis---AsnMetAspGlyAlaSerIleValIleGlnHisThrAsnGlyPhe 267
Db 819674 TTAGCGGGAATATTAAATGGTTATGTATATGTTATTTGTAAACAGCGCGAAACTGAT 819615
QY 268 ValSerSerTyrIleHisIleLysAspAlaGlnValIleThrGlyAspThrValArgThr 287
Db 819614 TTAAATTATATGGCTTCATCAAGCTGTATCACTGAATTTGGTCAGCTTTTCAGCA 819555
QY 288 GlyGlnArgIleAlaSerMetLysAsn-----GlnProSerGlyAlaAlaLeuPheGlu 305
Db 819554 GGGCAGGTTATTGCTCAATGAGAAATACAGGGGAATATCAGCTTCGCGCTTAT--- 819498
QY 306 PheArgIleSerArgAsnGlyValTyrValAspPro 317
Db 819497 TTGGTATTAGCCGTAAAGAACGCCAGTAATCCT 819462

RESULT 13
US-09-860-846-1/c
Sequence 1, Application US/09860846
Patent No. US20020164742A1
GENERAL INFORMATION:
APPLICANT: Sherman, D.H.
APPLICANT: Liu, H.
APPLICANT: Xue, Y.
APPLICANT: Zhao, L.
TITLE OF INVENTION: DNA encoding methymycin and pikromycin
FILE REFERENCE: 600,438US1
CURRENT APPLICATION NUMBER: US/09/860,846
PRIORITY FILING DATE: 2001-05-18
PRIOR APPLICATION NUMBER: 09/105,537
NUMBER OF SEQ ID NOS: 43
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 1
LENGTH: 15872
TYPE: DNA
ORGANISM: Streptomyces venezuelae

US-09-860-846-1

Alignment Scores:
Pred. No.: 0.00547 Length: 15872
Score: 130.50 Matches: 81
Percent Similarity: 38.75% Conservative: 43
Best Local Similarity: 25.31% Mismatches: 115
Query Match: 7.81% Indels: 81
DB: Gaps: 18

US-10-018-706-2 (1-322) x US-09-860-846-1 (1-15872)
QY 34 AlaSerLysProThrTyrAsnSerThrSerGlySerHisArgThrSerGlySer 53
Db 3195 GCGTCGCGCC-----TCGGCAGCAGCGCGCCACAGAGTCTCGGGGAG 3151
QY 54 -----GlyGlyLeuAlaIleGlySerGlnValIle 63
Db 3150 GTGACGCGCGCGGGGTAGCGCAGAGCCATCGCATGTATCGCATGGCGCGGTGTG 3091
QY 64 ThrAsp-----SerGlnGlyValProAsnArgTyrGlnValLysGlnGlyAspThrVal 81
```

```

Db      3090 TCCGCCGGGGTCCGGGGGATATCCCGGCGGATCCGCTCCGCCCGCGCGCGG 3031
Qy      82 SerLys-----IleAlaGlnArgTyrGlyLeuAsnTrpArglu----- 94
Db      3030 AGGAGGTGCCCGCAGTGGGCGGAGGCGCGCGCGGCTGGGTCGACAGACAGTCCG 2971
Qy      95 ILeGlyHisIleAsnAsnLeuAsnSerSerTyrThrIleTyrThrGlyInTrpLeuThr 114
Db      2970 CTGGGCGAGCGCGAGTCCCGCTGCTGTCGACGAGCGCGCTCCGCGAGC----- 2926
Qy      115 LeuTrpSerGlyAsp-----LeuIleValArgGluArgSer 126
Db      2925 -----TCGACGAGCATGAGGAGTGCAGCCAGCTCCTTGAAGGTGAGCGCGAGTTCG 2872
Qy      127 IleSerSerGly-----ValAsnThrAlaHisTrpProSerPro 139
Db      2871 ACCGCGGTCCGGTCTGCTACTCGAGGACGGCGCGCAGCTGTCGGCGACGCGCTCGTC 2812
Qy      140 ValAlaValAlaGlnSerSerArgProProValGlnGlnHisProAlaValGlnLysProThr 159
Db      2811 GTCTCGGCGAGCGCGCGCATCACC-----CCACCGCAGCCCGCGCCCTCG 2764
Qy      160 ProProValValValValLysLysProThrProThrProProValValGlnGlnProAla 179
Db      2763 CCC-----TCGCTCTGCGCTCTGCGCGCGAGCTCAGCTCAGACCCCGCGCGCGG 2722
Qy      180 ProValAlaProProValThrGluAlaProPheAlaThrGlySer----- 194
Db      2721 CCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2662
Qy      195 ---SerGlyValMetGlnPheArgTyrPro-----ValGlyAlaThrAsnProVal 211
Db      2661 CCGTTCGACCAAGTGGCGCTGCGCTGCGAGGCGGTAGTTCGCGAGCGACCCGACCCG 2602
Qy      212 ArgArgPheGlyThrAlaThrValAlaGlySer-----ThrValThr 225
Db      2601 CCGGTGCTCCCGTGGCGCGCGCTTCAGTCCAGTCCGTCGCGCGCGCGCGCGCGCG 2542
Qy      226 Ser-----AsnGlyMetTrpPheSerGlyArgAspGlyAspLeuIleAsnAlaSerAsn 243
Db      2541 AGTGGCGGAGCAGCAGCGCTGGCGGCTCGCGCGCGCGCGCGCGCGCGCGCGCG 2497
Qy      244 AlaGlyThrValIleGlnAlaAsp---HisAsnMetAspGlyAlaSerIleValIleGln 262
Db      2496 GCGGAGACCGCGCGTGGCGCGCTCTGCTGTCGCGTACGAGTCCGCGCATCGCGAGCAG 2437
Qy      263 HisThrAsnGlyPheValSerSerTyrIleHisIleLeuAsp---AlaGlnValLysThr 281
Db      2436 ACCCGCTCGGAGCGGAGCTCCAGGAAGGTTCGCGCGCGCGCGCGCGCGCGCGCG 2377
Qy      282 GlyAspThrValArgThrGlyGlnArgIleAlaSerMetLysAsnGlnProSerGlyAla 301
Db      2376 GCGTCGAGGAAGCTACGCGGCTCGCGG-----ACCTGCTCCACCCAGTACTCGGAGTGC 2323

```

```

; SEQ ID NO 1
; LENGTH: 15872
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-988-384B-1

Alignment Scores:
Pred. No.: 0.00547 Length: 15872
Score: 130.50 Matches: 81
Percent Similarity: 38.75% Conservative: 43
Best Local Similarity: 25.31% Mismatches: 115
Query Match: 7.81% Indels: 81
DB: 9 Gaps: 18

US-10-018-706-2 (1-322) x US-09-988-384B-1 (1-15872)

Qy      34 AlaSerLysProThrTyrAsnSerThrSerGlySerGlySerHisArgThrSerGlySer 53
Db      3195 GCGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3151
Qy      54 -----GlyGlyLeuAlaIleGlySerGlnValIle 63
Db      3150 GTGACCGCGCGCGGTAGCGCGCAGCCATGCCATGATCGCGATGCGGTGCGCGGTG 3091
Qy      64 ThrAsp-----SerGlnGlyValProAsnArgTyrGlnValLysGlnGlyAspThrVal 81
Db      3090 TCCGCCGGGGTCCGGGGGATATCCCGGCGGATCCCGGCTCCGCCCGCGCGG 3031
Qy      82 SerLys-----IleAlaGlnArgTyrGlyLeuAsnTrpArglu----- 94
Db      3030 AGGAGGTGCCCGCAGTGGGCGGAGGCGCGCGCGGCTGGGTCGACAGACAGTCCG 2971
Qy      95 ILeGlyHisIleAsnAsnLeuAsnSerSerTyrThrIleTyrThrGlyInTrpLeuThr 114
Db      2970 CTGGGCGAGCGCGAGTCCCGCTGCTGTCGACGAGCGCGCTCCGCGAGC----- 2926
Qy      115 LeuTrpSerGlyAsp-----LeuIleValArgGluArgSer 126
Db      2925 -----TCGACGAGCATGAGGAGTGCAGCCAGCTCCTTGAAGGTGAGCGCGAGTTCG 2872
Qy      127 IleSerSerGly-----ValAsnThrAlaHisTrpProSerPro 139
Db      2871 ACCGCGGTCCGGTCTGCTACTCGAGGACGGCGCGCAGCTGTCGGCGACGCGCTCGTC 2812
Qy      140 ValAlaValAlaGlnSerSerArgProProValGlnGlnHisProAlaValGlnLysProThr 159
Db      2811 GTCTCGGCGAGCGCGCGCATCACC-----CCACCGCAGCCCGCGCCCTCG 2764
Qy      160 ProProValValValValLysLysProThrProThrProProValValGlnGlnProAla 179
Db      2763 CCC-----TCGCTCTGCGCTCTGCGCGCGAGCTCAGCTCAGACCCCGCGCGCGG 2722
Qy      180 ProValAlaProProValThrGluAlaProPheAlaThrGlySer----- 194
Db      2721 CCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2662
Qy      195 ---SerGlyValMetGlnPheArgTyrPro-----ValGlyAlaThrAsnProVal 211
Db      2661 CCGTTCGACCAAGTGGCGCTGCGCTGCGAGGCGGTAGTTCGCGAGCGACCCGACCCG 2602
Qy      212 ArgArgPheGlyThrAlaThrValAlaGlySer-----ThrValThr 225
Db      2601 CCGGTGCTCCCGTGGCGCGCGCTTCAGTCCAGTCCGTCGCGCGCGCGCGCGCG 2542
Qy      226 Ser-----AsnGlyMetTrpPheSerGlyArgAspGlyAspLeuIleAsnAlaSerAsn 243
Db      2541 AGTGGCGGAGCAGCAGCGCTGGCGGCTCGCGCGCGCGCGCGCGCGCGCGCGCG 2497
Qy      244 AlaGlyThrValIleGlnAlaAsp---HisAsnMetAspGlyAlaSerIleValIleGln 262
Db      2496 GCGGAGACCGCGCGTGGCGCGCTCTGCTGTCGCGTACGAGTCCGCGCATCGCGAGCAG 2437
Qy      263 HisThrAsnGlyPheValSerSerTyrIleHisIleLeuAsp---AlaGlnValLysThr 281

```

```

Db      2436 ACCCGCGGACCGAGCTCAGAGAGGTGTCCGCGCCCGCATTCCTCCAGGGTCCGACG 2377
Qy      282 GtAspThrValArgThrGlyGlnArgIleAlaSerMetLysAsnGlnProSerGlyAla 301
Db      2376 GCGTCGAGAGGAGGTACGGGCGCTGCGG-----ACCTGCTCCACCGACTGCGGATCG 2323

RESULT 15
US-09-836-821-1/c
; Sequence 1, Application US/09836821
; Publication No. US2003087405A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.4380S1
; CURRENT APPLICATION NUMBER: US/09/836,821
; PRIOR FILING DATE: 2001-04-17
; PRIOR APPLICATION NUMBER: 09/105,537
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 15872
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-836-821-1

Alignment Scores:
Pred. No.:      0.00547      Length:      15872
Score:          130.50      Matches:      81
Percent Similarity: 38.75%      Conservative: 43
Best Local Similarity: 25.31%      Mismatches: 115
Query Match:      7.81%      Indels:      81
DB:               Gaps:      18

US-10-018-706-2 (1-322) x US-09-836-821-1 (1-15872)
Qy      34 AlaSerLysProThrTyraSerThrSerGlySerHisArgThrSerGlySer 53
Db      3195 GCGTCGCGGCGCC-----TCGGCGACGAGCGCGCCACAGGTCTCGGGGGAG 3151
Qy      54 -----GlyGlyLeuAlaIleGlySerGlnValIle 63
Db      3150 GTGACCGCGCGGGGTAGCGGAGCCGATGCCATGATCGCGATCGCGGTGG 3091
Qy      64 ThrAsp-----SerGlnGlyValProAsnArgTyrcInValLysGlnGlyAspThrVal 81
Db      3090 TCCGCGCGGGGTCCGGCGGATATCCGTCGCGCGGATCCGATCGCGCGGTG 3031
Qy      82 SerLys-----IleAlaGlnArgTyrcLysAsnTyrcArgLys 94
Db      3030 AGCAGGTCCGCCAGGTGGGCGGAGGCGCGCGCGTGGGTGTCAGAGACAGATCCG 2971
Qy      95 IleGlyHisIleAsnLeuAsnLeuAsnSerTyrcIleTyrcGlyGlnTyrcPleuThr 114
Db      2970 CTGGGCGAGCGGCGCGCGTGTGTCGACGAGCGCGTTCGCGAGC-----2926
Qy      115 LeuTyrcSerGlyAsp-----LeuLysValArgGlyLysSer 126
Db      2925 -----TCACGCGACATGAGGAGTGAAGCCAGCTCTTGAAGTGAAGCCGAGTTCG 2872
Qy      127 IleSerSerGly-----ValAsnThrAlaHisThrProSerPro 139
Db      2871 ACCGCGGTGGGTGTCGTCTACTGACGAGACGCGCGCGATGTCGCGCGACGCGCTCGTGC 2812
Qy      140 ValAlaValGlnSerSerArgProProValGlnGlnHisProAlaValGlnLysProThr 159
Db      2811 GTCTCGTGGGAGCGCGCGATGACCG-----CCACCGCGACCGCGCGCTTCG 2764
Qy      160 ProProValValValLysLysProThrProThrProProValValGlnGlnProAla 179

```

```

Db      2763 CCC-----TCGCTCGCGCTCGCGCCGACGTCACACCGCGCGCGG 2722
Qy      180 ProValAlaProProValThrGlnAlaProPheAlaThrGlySer-----194
Db      2721 CCGGTGCCCGACCGCGCGGTCCGATTCGCGCCGCGCGCGCGCGCGCTTCGCGCGCG 2662
Qy      195 ---SerGlyValMetGlnPheArgTyrcPro-----ValGlyAlaThrAsnProValVal 211
Db      2661 CCGTCCGACCAAGCGCGCGCTTCGCTGCGTGAAGCGTGAAGTCCGACGCGGACCCCTACCGTGC 2602
Qy      212 ArgArgPheGlyThrAlaThrValAlaGlySer-----ThrValThr 225
Db      2601 CCGGTGCTCCCGGCGCGCGCGCTCCAGTCCAGTCGCGCGCGCGCGCGCGCGGTG 2542
Qy      226 Ser-----AsnGlyMetTyrcPheSerGlyArgAspGlyAspLeuIleAsnAlaSerAsn 243
Db      2541 AGTCCGCGCGAGCGACGCGATCGGCTCCGCGCGCGCG-----TTGCGGACG 2497
Qy      244 AlaGlyThrValIleGlnAlaAsp---HisAsnMetAspGlyAlaSerIleValIleGln 262
Db      2496 GCGGAGACCGCGCGCGCTCGCTCGTCCGTCGCGTCCGATCGGAGTCCGCGCATCGCGGACG 2437
Qy      263 HisThrAsnGlyPheValSerSerTyrcIleHisIleLysAsp---AlaGlnValLysThr 281
Db      2436 ACCCGTCGCGGACCGAGCTCCAGAGAGGTGTCCGCGCGCGCATTCCTCCAGGGTCCGTAAG 2377
Qy      282 GtAspThrValArgThrGlyGlnArgIleAlaSerMetLysAsnGlnProSerGlyAla 301
Db      2376 GCGTCGAGAGGAGGTACGGGCGCTGCGG-----ACCTGCTCCACCGACTGCGGATCG 2323

RESULT 16
US-09-861-289-1/c
; Sequence 1, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.4380S1
; CURRENT APPLICATION NUMBER: US/09/861,289
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 15872
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-1

Alignment Scores:
Pred. No.:      0.00547      Length:      15872
Score:          130.50      Matches:      81
Percent Similarity: 38.75%      Conservative: 43
Best Local Similarity: 25.31%      Mismatches: 115
Query Match:      7.81%      Indels:      81
DB:               Gaps:      18

US-10-018-706-2 (1-322) x US-09-861-289-1 (1-15872)
Qy      34 AlaSerLysProThrTyraSerThrSerGlySerHisArgThrSerGlySer 53
Db      3195 GCGTCGCGGCGCC-----TCGGCGACGAGCGCGCCACAGGTCTCGGGGGAG 3151
Qy      54 -----GlyGlyLeuAlaIleGlySerGlnValIle 63
Db      3150 GTGACCGCGCGGGGTAGCGGAGCCGATGCCATGATCGCGATCGCGGTGG 3091
Qy      64 ThrAsp-----SerGlnGlyValProAsnArgTyrcInValLysGlnGlyAspThrVal 81

```

Wed Jul 9 10:01:00 2003

us-10-018-706-2.rnpb

Page 14

```
Db 3090 TCCGCCGGGGTGGCGGGCGGATCCCTGCTCCGCCGCTGCCCGCGGTG 3031
QY 82 SerIys-----IleAlaGlnArgIYrGlyLeuAsnTPyrArgIu----- 94
Db 3030 AGCAGGTCGCCCAAGTGGCGGCGAGCGCGCGGCGGTGGTGGAGGAGCATCCG 2971
QY 95 ILeGlyHisIleAsnMetLeuAsnSerIYrThrIleYrThrGlyGlnTPreunThr 114
Db 2970 CTGGCGCAGGCGGATCCCGTGTCTGACAGAGCGGCTCCGAGC----- 2926
QY 115 LeuTPSerGlyAsp-----LeuIysValIArgIuArgSer 126
Db 2925 -----TCGACGGAATAGAGGAGTCCAGAACCCAGCTCTTGAAGTGGAGCCGAGTTCCG 2872
QY 127 ILeSerSerGly-----ValAsnThrAlaHisThrProSerPro 139
Db 2871 AGCGCGGTGGGTCTGCTGCTGACCTCGAGAGCGCGCGGAGCTGTGGCGCAGCGCTCGCTC 2812
QY 140 ValAlaValGlnSerSerIYrProProValGlnGlnHisProAlaValGlnIYrProThr 159
Db 2811 GTCTGTGGCGGAGCGCGCGGATCAACG-----CCACCCGACCGCGCGCTCG 2764
QY 160 ProProValIValIValIYsPProThrProThrProProValIValGlnIYrProAla 179
Db 2763 CCC-----TCGCGCTCGCGCTCGCGCGGACGTCACACCGCGCGCGCGG 2722
QY 180 ProValAlaProProValIThrGlnAlaProPheAlaThrGlySer----- 194
Db 2721 CCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2662
QY 195 -----SerGlyValMetGlnPheArgIYrPro-----ValGlyAlaThrAsnProValI 211
Db 2661 CCGTGAACACGAGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2602
QY 212 ArgArgPheGlyThrAlaThrValAlaGlySer-----ThrValIThr 225
Db 2601 CCGGTGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2542
QY 226 Ser-----AsnGlyMetIYrPheSerGlyArgAspGlyAspLeuIleAsnIleAsn 243
Db 2541 AGTGGCGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2497
QY 244 AlaGlyThrValIleGlnIleAsnAsp---HisAsnMetAspGlyAlaSerIleValIleGln 262
Db 2496 GCGGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2437
QY 263 HisThrAsnGlyPheValSerSerIYrIleHisIleIYsAsp---AlaGlnValIYrThr 281
Db 2436 ACCCGGCGGAGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2377
QY 282 GlyAspThrValIArgIYrGlnIYrGlnIleAsnSerMetIYsAsnGlnProSerGlyAla 301
Db 2376 GCGTGGAGAGGAGCGTACCGGCGCTGCG-----ACCTGTCTACCGCGATCTCGGATCG 2323
RESULT 17
US-10-087-667-4
Sequence 4, Application US/10087667
Publication No. US20020194629A1
GENERAL INFORMATION:
APPLICANT: Bramley, John A.
APPLICANT: Plaut, Karen I.
TITLE OF INVENTION: TREATMENT OF STAPHYLOCOCCUS INFECTIONS
FILE REFERENCE: Mastitis
CURRENT FILING DATE: 2002-02-28
NUMBER OF SEQ ID NOS: 10
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 4
LENGTH: 1520
TYPE: DNA
ORGANISM: Staphylococcus simulans
```

```
US-10-087-667-4
Alignment Scores:
Pred. No.: 0.000338
Score: 128.50
Percent Similarity: 33.98%
Best Local Similarity: 22.01%
Query Match: 7.69%
DB: 9
Gaps: 15
US-10-018-706-2 (1-322) x US-10-087-667-4 (1-1520)
QY 10 GlnAsnGlnIYrProIleIYsArgIYrGlnIlePheGlyValIleThrThrCysIle 29
Db 313 CAGCATATGAGAGAGATTTCCAGAGCGGAGCTG-----GGCTGGCGCGTGTGTGCGG 366
QY 30 LeuAlaGlyCysAlaSerIYrProThrIYrAsnSerThrSerIYrSerGlySerHisArg 49
Db 367 CTGGCG-----ACGATCGCGCGCAACGCGCGCGAGG 399
QY 50 ThrSerGlySerGlyIYrLeuAlaIleGlySerGlnValIleThrAspSerGlnIYrVal 69
Db 400 GCCACGGGCTCAG-----CGCGAGGATCTGTGTATTTCTACGAGATG----- 444
QY 70 ProAsnArgIYrGlnIYrValIYsGlnIYrAspThrValSerIYrIleAlaGlnArgIYrGly 89
Db 445 -----TTGCACTTCGACATCGATCGCATCTGCGCAGCATCGCGCATCTGCAC 495
QY 90 LeuAsnTPyrArgIuIleGlyHisIleAsnMetLeuAsnSerSerIYrThrIleYrThr 109
Db 496 AAGCACTCGGAGAGATCTCGCAC----- 519
QY 110 GlyIYrIlePheThrIleuTPSerGlyAspLeuIYsValIArgIuArgSerIleSerSer 129
Db 520 -----TGGCGCGC-----TACAGCGGATTCAGCGCA 546
QY 130 GlyValAsnThrAlaHisThrProSerProValAlaValGlnSerSerIYrProProVal 149
Db 547 AGTGTATCGCGCTGATGAGACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 606
QY 150 GlnGlnHisProAlaValGlnIYrProThrProProValIValIYrIYsPro--- 168
Db 607 GTCCGTGGCAAGCTGCGCGCGCGCGCGCG-----GCTTCGCGCGCGAGACCGCG 657
QY 169 -----ThrProThrProProValIValGln----- 176
Db 658 AGTTCGCGTGGCGCTGCGCGAGTCTGTACAGCGGATCCGACGCGCGCAAGGGGCC 717
QY 176 ----- 176
Db 718 GGTACGCTGGCGCGCGCGCAATCCGCTGCAGCGCGCTTCGAGCGTTCCGCGCAACGA 777
QY 177 -----GlnProAlaProValAlaProProValIThr 186
Db 778 GCCGCGCGCGCGCTGCGCGCGCGAGGATTCAGCTGTACAGCGCGCTGTTCGA 837
QY 187 Gln---AlaPro-----Phe 190
Db 838 CGAACCGGCGGAGCGGAGCGGCTTCGAGCGCGCTTCGCAAGCGCGCGCGCAAGCTGCA 897
QY 191 AlaThrIYrSerSerGlyValMetGlnPheArgIYrProValIYrAlaHisAsnProVal 210
Db 898 GCGGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 957
QY 211 ValArgArgPheGlyThrAlaThrValAlaGlySer----- 222
Db 958 -----GGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 222
QY 223 ThrValIYrSerAsnGlyMetIYrPheSerGlyArgAspGlyAspLeuIleAsnIleSer 242
Db 1009 GACATGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1068
QY 243 AsnAlaGlyThrValIleGlnIleAsnHisAsnMetAspGlyAlaSerIleValIleGln 262
```

```
Db 1069 GCCCGCGCTGCTTCAACGCG-----CACTTGTGCTTGGCGGAGATCGTG----- 1116
Qy 263 HishtrAnsglyPheValSerSerTyrlleHsiIlelyAspAlaGlnVallyThrGly 282
Db 1117 CACACCGCGGCTGCTGCACGACCACTTACCACTGATGAACCTTACCAACCGCG 1176
Qy 283 AsphThrValArgThrGlnGlyAlaArgIleAlaSerMetlyAsnGlnProSerGlyAla 301
Db 1177 GCCACAGTGTGATGAACACCGCCATCGCCCAACCGCCCAACCCAGGCGAGCGC 1233

RESULT 18
US-09-468-147-164
; Sequence 164, Application US/09468147A
; Publication No. US20030049601A1
; GENERAL INFORMATION:
; APPLICANT: Abbott Laboratories
; APPLICANT: Schlander, George G.
; APPLICANT: Erker, James C.
; APPLICANT: Desai, Suresh M.
; APPLICANT: Dawson, George J.
; APPLICANT: Mushawar, I. K.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTING
; TITLE OF INVENTION: HEPATITIS E VIRUS
; FILE REFERENCE: 6232.US.PI
; CURRENT APPLICATION NUMBER: US/09/468,147A
; EARLIER FILING DATE: 1999-12-21
; EARLIER APPLICATION NUMBER: US 09/173,141
; EARLIER FILING DATE: 1998-10-15
; EARLIER APPLICATION NUMBER: US 60/061,199
; NUMBER OF SEQ ID NOS: 258
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 164
; LENGTH: 7277
; TYPE: DNA
; ORGANISM: Hepatitis E Virus
; FEATURE:
; OTHER INFORMATION: us2full1
US-09-468-147-164

Alignment Scores:
Pred. No.: 0.00438 Length: 7277
Score: 127.00 Matches: 80
Percent Similarity: 33.14% Conservative: 37
Best Local Similarity: 22.66% Mismatches: 120
Query Match: 7.60% Indels: 116
DB: 9 Gaps: 19

US-10-018-706-2 (1-322) x US-09-468-147-164 (1-7277)
Qy 27 ThrCysIleLeuAlaGlyCys-----AlaSerLysProThrTyraSerSerThrSer 43
Db 1755 ACGAGGAGTGTGAATGAGCGCCATCTTGAAGCGAATGCGCCCTGAGAGTATGTTCTGCA 1814
Qy 44 GlySerGlySerHisArgThrSerGlySerGlyGlyLeuAlaIle----- 58
Db 1815 TTGAGCGCTCTCGCGAGCTTATGAGGCGCGGCGGTCGACAGCCTCACTTATGAGTCAACC 1874
Qy 59 -----GlySerGlnValIleThrAspSerGlnGlyValProAsnArgTyrglnValLys 76
Db 1875 CCGCGCGGTCTGCGAGTAAAGTTTCATCTATATGTTGATGATGACATGCCCATTTCCC 1934
Qy 77 GlnGlyAspThrValSer----- 82
Db 1935 CCYGGTGGCGCCCTTAGCGCGCGCGGAGAGTGGCGGCTTCTGACAGTCTTTAT 1994
Qy 83 -----LysIleAlaGlnArgTyrglyLeuAsn-----TryArg----- 93
Db 1995 AGATCAATAGTGTACCCAGCGGCAATTCGCTGACAGCGGAGCTATAGCTATCTCTAG 2054
Qy 94 -----GlnIleGlyHisIle-----AsnAsn 100
```

```
Db 2055 GAGCGTCTGGGATACCTTCCCCCATTTCTCCCTGGGCGATATTGGAGATCTGTACCCC 2114
Qy 101 LeuAsnSerSerTyrlleTyrlleThrGlyGlnTrpLeuThrLeu---TrpSerGlyAsp 119
Db 2115 TTTTGGCGTGAAGGAGCTTGTATATCCGAACTGGCAACCTCTGTTTCTAGAT 2174
Qy 120 LeuIleValArgIleArgSerIleSerSerGlyValAsnThrAlaHisThrProSerPro 139
Db 2175 TTTCTCCCCCTTAGAGGGCGGCGCTCTGCTTGGCTGCGCCCGCGGGGTTGCCCTACCT 2234
Qy 140 ValAlaValGlnSerSerArgProVal----- 149
Db 2235 ACT-----CCACCTGTATGATATCTGGGTATTACACCGCCCTCA 2276
Qy 150 -----GlnGlnHis-----ProAlaValGlnLysPro-----Thr 159
Db 2277 GAGGAATCTCATGTGATGAGGCACTGTACCTCTCTCTCTGAGCCCTGCTGATTAAC 2336
Qy 160 ProProValAlaValAlaValLysProThrProThrProProValAlaGlnGlnProAla 179
Db 2337 AGCCCTATTGTGCTTACC-----CCCCCCCCCCCCCTCTCCGTTGCGTAACCGGCA 2390
Qy 180 ProValAlaProProValThrGlnAlaProPheAlaThrGlySerSerGlyValMetGln 199
Db 2391 ACATCCCCCGCTCCCGCCACTCGCCCT-----CTCCTT 2423
Qy 200 PheArgTyrlleProValGlyAlaThrAsnProValAlaArgArgPheGlyThrAlaThrVal 219
Db 2424 TACACCTTACCCCGACGCGCGC-----AACGTGAT 2453
Qy 220 AlaGlySerThrValThrSerAsnGlyMetTrpPheSerGlyArgAspGlyAspLeuIle 239
Db 2454 GCGGGGTCACTTGTGTAGTCAAGTGTGATG-----TTAGTC 2492
Qy 240 AsnAlaSerAsnAlaGlyThrValIleGlnAlaAspHisAsnMetAspGlyAlaSerIle 259
Db 2493 AATGCTCTCAAAACCTGTGC-----CATGCGCCCGGGGGTGC----- 2528
Qy 260 ValIleGlnHisThrAnsglyPheValSerSerTyrlleHsiIlelyAspAlaGlnVal 279
Db 2529 ---CTGTGCAT-----GCTTTTATCAAGCTTCCAGAGCGCTTCTAC 2570
Qy 280 LysThrGlyAspThrValArgThrGlyGlnArgIleAlaSerMetlyAsnGlnPro--- 298
Db 2571 TCGACTGAATTCATCATCGCGGAGGCGCTTGCACATACCTTAAACCCCGCCCTATT 2630
Qy 299 SerGlyAlaAlaLeuPheGlnPheArgIleSerArgAsn 311
Db 2631 ATCCATGCAATGCTCCCGACTATATGAGTTGAGCAAAAC 2669

RESULT 19
US-09-468-147-165
; Sequence 165, Application US/09468147A
; Publication No. US20030049601A1
; GENERAL INFORMATION:
; APPLICANT: Abbott Laboratories
; APPLICANT: Schlander, George G.
; APPLICANT: Erker, James C.
; APPLICANT: Desai, Suresh M.
; APPLICANT: Dawson, George J.
; APPLICANT: Mushawar, I. K.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTING
; TITLE OF INVENTION: HEPATITIS E VIRUS
; FILE REFERENCE: 6232.US.PI
; CURRENT APPLICATION NUMBER: US/09/468,147A
; EARLIER FILING DATE: 1999-12-21
; EARLIER APPLICATION NUMBER: US 09/173,141
; EARLIER FILING DATE: 1998-10-15
; EARLIER APPLICATION NUMBER: US 60/061,199
; NUMBER OF SEQ ID NOS: 258
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 165
```



```

; TITLE OF INVENTION: HEPATITIS E VIRUS
; FILE REFERENCE: 6232 US.P1
; CURRENT APPLICATION NUMBER: US/09/468,147A
; EARLIER FILING DATE: 1999-12-21
; EARLIER APPLICATION NUMBER: US 09/173,141
; EARLIER FILING DATE: 1998-10-15
; EARLIER APPLICATION NUMBER: US 60/061,199
; EARLIER FILING DATE: 1997-10-15
; NUMBER OF SEQ ID NOS: 258
; SOFTWARE: FASCSeg for Windows Version 3.0
; SEQ ID NO 141
; LENGTH: 903
; TYPE: DNA
; ORGANISM: Hepatitis E Virus
; OTHER INFORMATION: us2-851
; US-09-468-147-141

Alignment Scores:
Pred. No.: 0.000388 Length: 903
Score: 125.00 Matches: 61
Percent Similarity: 35.95% Conservative: 26
Best Local Similarity: 25.21% Mismatches: 81
Query Match: 7.49% Indels: 74
DB: 9 Gaps: 14

US-10-018-706-2 (1-322) x US-09-468-147-141 (1-903)
QY 96 GlyHisLe-----AsnAsnLeuAsnSerSerTyrrHlleTyrrHlyGln 111
Db 205 GGGCTATTGTTGGAGAGTGTCTAACCCCTTTGGCGGAGGAGGACTTGTATCCAGACC 264
QY 112 TrpLeuThrLeu---TrpSerGlyAspLeuValArgGlnArgSerIleSerSerGly 130
Db 265 TGGTCAAGCTCTGCTTTTCTAGTGAATTTCTCCCTAGAGCGCGCTCTCTGCTTGG 324
QY 131 ValAsnThrAlaHisThrProSerProValAlaValGlnSerSerArgProProVal 149
Db 325 GCTGCGCGCGCGGAGTGTGCTTACCTTACT-----CCACCTGTACT 366
QY 150 -----GlnGlnHis-----GlnGlnHis-----Pro 153
Db 367 GATATCTGGGTGTACACCGCCCTCAGAGAACTCATGTTGANTGCGCATCTGACCC 426
QY 154 AlaValGlnLysPro-----ThrProValAlaValLysLysProThrPro 170
Db 427 TCTGTTCTGAGCGCTGCTGATTGACACGCGCTATTGTGCTTACC-----CCCCCCCC 480
QY 171 ThrProValAlaGlnGlnProAlaProValAlaProProValThrGlnAlaProHe 190
Db 481 CCCCTCTCTCCGCTGAGTAAAGCGGACATCCCGCTCTCCCGACCTGCGCGT----- 534
QY 191 AlaThrGlySerSerGlyValMetGlnPheArgTyrrProValGlyAlaThrAsnProVal 210
Db 535 -----CTCTTACACTTACCCGACGCGCC----- 561
QY 211 ValArgArgPheGlyThrAlaThrValAlaGlySerThrValThrSerAsnLysMetTrp 230
Db 562 -----AAGGTGTATGCGGGGCTCTGTGTGAGTCAACATGTGATGTG 603
QY 231 PheSerGlyArgAspGlyAspLeuIleAsnAlaSerAsnAlaGlyThrValIleGlnAla 250
Db 604 -----TTAGTCAATGCTCTCAACCTTGGC----- 627
QY 251 AspHisAsnMetAspGlyAlaSerIleValIleGlnHisThrAsnGlyPheValSerSer 270
Db 628 -----CATCGCGCGCGGGTGGC-----CTCTGCGCAT-----GCTTTT 660
QY 271 TyrrIleHisIleLysAspAlaGlnValLysThrGlyAspThrValArgThrGlyGlnArg 290
Db 661 TATTCACGTTTCCCAACAGCTTCTACTGATGATTCATGATGCGGAGGCGCTTGA 720
QY 291 IleAlaSerMetLysAsnGlnPro---SerGlyAlaIleLeuPheGlnPheArgIleSer 309
```

```

Db 721 GCATACACTTTTAAACCCCGCCCTTATTCATCATGATGCTCCCGACTATAGGTGAG 780
QY 310 ArgAsn 311
Db 781 CAAAC 786

RESULT 23
US-10-063-547-99
; Sequence 99, Application US/10063547
; Publication No. US20020182638A1
; GENERAL INFORMATION:
; APPLICANT: Baton, Dan L.
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Marabe, Colin K.
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3230P1C1
; CURRENT APPLICATION NUMBER: US/10/063,547
; PRIOR APPLICATION removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 170
; SEQ ID NO 99
; LENGTH: 2436
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-063-547-99

Alignment Scores:
Pred. No.: 0.00289 Length: 2436
Score: 122.50 Matches: 78
Percent Similarity: 39.04% Conservative: 52
Best Local Similarity: 23.42% Mismatches: 145
Query Match: 7.34% Indels: 60
DB: 9 Gaps: 12

US-10-018-706-2 (1-322) x US-10-063-547-99 (1-2436)
QY 6 AlaIleAsnSerGlnAsnGlnLysProIleLysArgLeuGlyLeuIlePheGlyValIle 25
Db 686 GCCACCAACTTGTAGTCCAGACAGTGTCCAGTGGGCCAGCACTGCCACCACTGTGAG 745
QY 26 ThrThrCysIleLeuAlaGlyCysAlaSerLysProThrTyrrAsnSerThrSerGlySer 45
Db 746 TCTAGCACACTCTCCAGTGGGCGC-----AGCAGCGACCACTCT 787
QY 46 GlySerHisArgThrSerGlySerGlyGlyLeuAlaIleGlySerGlnValIleThrAsp 65
Db 788 GACTTCAGCACACACTTCCAGTGGGCTAGACACGCCCAACTCTGTGATCCAGACAAACC 847
QY 66 SerGlnGlyValProAsnArgTyrrGln-----ValLysGlnGlyAsp 79
Db 848 TCCAGTGGGGCGCAG---CAGCAGCAGCACTGTGATCCAGACAGTGTCCAGTGGGCGCAG 906
QY 80 ThrValSerLysIleAlaGlnArgTyrrGlyLeuAsnTrpArgGlnIleGlnHis----- 97
Db 907 CACTGCCACCAACTCTGAGTCCAGACCACTCTCAGTGGGCGCAGCAGCAGCACTC 966
QY 98 -----IleAsnAsnLeu---AsnSerSerTyrrHlleTyrrThrGlyGlnTrpLeuThrIle 115
Db 967 TGAATCCAGAAACGACTCTCAATGCGCTGCGACACCACTCTGATCCAGACGAC 1026
QY 115 utrSerGlyAspLeuLysValArg-----GluArgSerIleSerSerGlyValAs 132
Db 1027 CTCAGTGGGGCGCAGCAGCAGCAGCAACTGTGATCCAGACAGTGTCCAGTGGGCGCAG 1086
QY 132 nThrAla-HisThrProSerProValAlaValGlnSerSerArgProProVal-----G 150
```

```

Db      1087 CACTGCCACCACTGTGAGTCCA-----GCACGACCTCCAGTGGGGCCAG 1131
Qy      150 IngInhiSProAlaValGlnInysProThrProProValValValyslySerProThrP 170
Db      1132 CACAGCCACCACTGTGAGTCCAGACGACGAGTGGGGCTGAGCCAGCACGACCACTC 1191
Qy      170 roThrProProValValGlnInPro--AlaProValAlaProProVal-----T 186
Db      1192 TGACTCCA-----GCACACTCTCCAGTGGGGCCGACAGCCACCACTGTGAGTCCAG 1245
Qy      186 hrGlnAlaProPheAlaThrGlySerSerGlyValMetGlnPheArgTyrProValGlyA 206
Db      1246 CACAGTGTCCAGTGGGATCAGACAGTCACCAATTCTGAGTCCAGACACCTCC-AGTG 1304
Qy      206 laThrAspProValValArgArgPheGlyThrAlaThrValAlaGlySerThrValThrs 226
Db      1305 GGGCCAC------ACAGCCACCACTGTGAGTCCAGTCCAGTCCAGTCC 1343
Qy      226 erAsnGlyMetTrpPheSerGlyArgAspGlyAspLeuIleAsnAlaSerAsnAlaGlyT 246
Db      1344 CCAGTGG------GCCAACACAGCCA 1364
Qy      246 hrValIleGlnAlaAspHisAsnMetAspGlyAlaSerIleValIleGlnHis----- 263
Db      1365 CCAACTGTGAGTCCAGACAGAGTGTCCAGTGGGGCCAGACAGTCCACCAACTGTGAGTCCA 1424
Qy      264 -----ThraSnglyPheValSerSerTyrIleHisIleLysAspAlaGlnValysThrG 282
Db      1425 GCACAACTCCAGTGGGGTCCAGACAGCCACCACTGTGAGTCCAGCAACTCCAGTCC 1484
Qy      282 lyaSPThrValArgThrGlyGlnArgIleAlaSerMetLysAsnGlnProSerGlyAlaA 302
Db      1485 GGGGTAGACAGACCACTGTGAGTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGTCC 1544
Qy      302 lalaupheGlnPheArgIleSerArgAsnGlyVal 313
Db      1545 CCAACTGTGAGTCCAGACAGTGTCCAGTGGGATC 1579

```

RESULT 24

```

US-10-174-590-309
: Sequence 309, Application US/10174590
: Publication No. US20030008352A1
: GENERAL INFORMATION:
: APPLICANT: Baker, Kevin P.
: APPLICANT: Chen, Dian
: APPLICANT: Desnoyers, Luc
: APPLICANT: Goddard, Audrey
: APPLICANT: Godowski, Paul J.
: APPLICANT: Gurney, Austin L.
: APPLICANT: Pan, James
: APPLICANT: Smith, Victoria
: APPLICANT: Watanabe, Colin K.
: APPLICANT: Wood, William I.
: APPLICANT: Zhang, Zemin
: TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
: FILE REFERENCE: P3430P1CA2
: CURRENT FILING DATE: 2002-06-18
: Prior application removed - See File Wrapper or Palm
: NUMBER OF SEQ ID NOS: 612
: SEQ ID NO 309
: LENGTH: 2436
: TYPE: DNA
: ORGANISM: Homo Sapien
US-10-174-590-309

```

```

Alignment Scores:
Pred. No.: 0.00289
Score: 122.50
Percent Similarity: 39.04%
Best Local Similarity: 23.42%
Length: 2436
Matches: 78
Conservative: 52
Mismatch: 145

```

```

Query Match: 7.34% Indels: 60
DB: 9 Gaps: 12
US-10-018-706-2 (1-322) x US-10-174-590-309 (1-2436)
Qy      6 AlaIleAsnSerGlnAsnGlnLysProIleLysArgLeuGlyLeuIlePheGlyValIle 25
Db      686 GCCACCAACTGTGAGTCCAGACAGTGTCCAGTGGGGCCAGACCTGACCACTGTGAG 745
Qy      26 ThrThrCysIleLeuAlaGlyCysAlaSerLysProThrTyrTrpAsnSerThrSerGlySer 45
Db      746 TCTAGCACTCTCCAGTGGGCC-----AGCACGCCACCACTCT 787
Qy      46 GlySerHisArgThrSerGlySerGlyValLeuAlaIleGlySerValIleThrAsp 65
Db      788 GACTCCAGACCACTCTCAGTGGGGCTAGACAGCCACCACTGTGAGTCCAGACCAACC 847
Qy      66 SerGlnGlyValProAsnArgTyrGln-----VallySnglyAsp 79
Db      848 TCCAGTGGGGCCAG-CACAGCCACCACTGTGAGTCCAGACAGTGTCCAGTGGGGCCAG 906
Qy      80 ThrValSerLysIleAlaGlnArgTyrGlyLeuAsnTrpArgGluIleGlyHis----- 97
Db      907 CACTGCCACCACTGTGAGTCCAGACCACTCCAGTGGGGCCAGACCACTC 966
Qy      98 -----IleAsnLeu-AsnSerSerTyrThrIleTyrThrGlyGlnTrpLeuThrle 115
Db      967 TGAGTCCAGAAACGACTCTCAATGGGGCTGGACAGCCACCACTGTGAGTCCAGACGAC 1026
Qy      115 utrSerGlyAspLeuLysValArg-----GluArgSerIleSerSerGlyValAs 132
Db      1027 CTCAGTGGGGCCAGACAGCCACCACTGTGAGTCCAGACAGTGTCCAGTGGGGCCAG 1086
Qy      132 ntThrAla-HisThrProSerProValAlaValGlnSerArgProProVal-----G 150
Db      1087 CACTGCCACCACTGTGAGTCCA-----GCACGACCTCCAGTGGGGCCAG 1131
Qy      150 IngInhiSProAlaValGlnLysProThrProProValValValyslySerProThrP 170
Db      1132 CACAGCCACCACTGTGAGTCCAGACGAGTGTCCAGTGGGGCTGAGCCAGCACCACTC 1191
Qy      170 roThrProProValValGlnInPro--AlaProValAlaProProVal-----T 186
Db      1192 TGACTCCA-----GCACACTCTCCAGTGGGGCCGACAGCCACCACTGTGAGTCCAG 1245
Qy      186 hrGlnAlaProPheAlaThrGlySerSerGlyValMetGlnPheArgTyrProValGlyA 206
Db      1246 CACAGTGTCCAGTGGGATCAGACAGTCACCAATTCTGAGTCCAGACACCTCC-AGTG 1304
Qy      206 laThrAspProValValArgArgPheGlyThrAlaThrValAlaGlySerThrValThrs 226
Db      1305 GGGCCAC------ACAGCCACCACTGTGAGTCCAGTCCAGTCCAGTCCAGTCC 1343
Qy      226 erAsnGlyMetTrpPheSerGlyArgAspGlyAspLeuIleAsnAlaSerAsnAlaGlyT 246
Db      1344 CCAGTGG-----GCCAACACAGCCA 1364
Qy      246 hrValIleGlnAlaAspHisAsnMetAspGlyAlaSerIleValIleGlnHis----- 263
Db      1365 CCAACTGTGAGTCCAGACAGTGTCCAGTGGGGCCAGACAGTCCACCAACTGTGAGTCCA 1424
Qy      264 -----ThraSnglyPheValSerSerTyrIleHisIleLysAspAlaGlnValysThrG 282
Db      1425 GCACAACTCCAGTGGGGTCCAGACAGCCACCACTGTGAGTCCAGCAACTCCAGTCC 1484
Qy      282 lyaSPThrValArgThrGlyGlnArgIleAlaSerMetLysAsnGlnProSerGlyAlaA 302
Db      1485 GGGGTAGACAGACCACTGTGAGTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGTCC 1544
Qy      302 lalaupheGlnPheArgIleSerArgAsnGlyVal 313
Db      1545 CCAACTGTGAGTCCAGACAGTGTCCAGTGGGATC 1579

```

```
RESULT 25
US-10-176-758-309
; Sequence 309, Application US/10176758
; Publication No. US2003008353A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430P1C104
; CURRENT APPLICATION NUMBER: US/10/176,758
; CURRENT FILING DATE: 2002-06-21
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 309
; LENGTH: 2436
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-176-758-309

Alignment Scores:
Pred. No.: 0.00289 Length: 2436
Score: 122.50 Matches: 78
Percent Similarity: 39.04% Conservative: 52
Best Local Similarity: 23.42% Mismatches: 145
Query Match: 7.34% Indels: 60
DB: Gaps: 12

US-10-018-706-2 (1-322) x US-10-176-758-309 (1-2436)
QY 6 AlaIleAsnSerGlnAsnGlnIysProIleuYsArGleuGlyLeuIlePheGlyValIle 25
DB 686 GCACCAACTCTGAGTCCAGCAGACAGTCCAGTGGGCGACAGCTGCCACCAACTCTGAG 745
QY 26 ThrThrCysIleLeuAlaGlyCysAlaSerIysProThrIYrAsnSerThSergIYser 45
DB 746 TCTAGCACCTCTCCAGTGGGCC-----AGCACGCCACCACTCT 787
QY 46 GlySerHisArgThSergIYserGlyLeuAlaIleGlySergIYValIleThrAsp 65
DB 788 GACTCCAGCAGACCACTCTGAGTGGGCTAGCAGACCACTCTGAGTCCAGCAGAAC 847
QY 66 SerGlnGlyValProAsnArgTrgIn-----VallyGlnGlyAsp 79
DB 848 TCCAGTGGGCGCAG-CACAGCCACCACTCTGAGTCCAGCAGACGTGTCAGTGGGCGCAG 906
QY 80 ThrValSerIYleAlaGlnArgTrgIYleuAsnTrpArgIYleGlyHis----- 97
DB 907 CACTGCCACCACTCTGAGTCCAGCAGCACTCTGAGTGGGCGCAGCAGCAGCAGCACTC 966
QY 98 -----IleAsnAsnLeu-AsnSerSerTrgThrIleTrgIYleuInTrpLeuThrIle 115
DB 967 TGAGTCCAGACAGACTCTCCATGAGGCTGGCAGACCACTCTGAGTCCAGCAGCAGCAG 1026
QY 115 utrPserGlyAspLeuIYleValArg-----GluArgSerIleSerSergIYValAs 132
DB 1027 CTCACAGTGGGCGCAGCAGCAGCAGCACTCTGAGTCCAGCAGACAGTCCAGTGGGCGCAG 1086
QY 132 nThrAla-HisThrProSerProValAlaValGlnSerSerTrgProProVal-----G 150
DB 1087 CACTGCCACCACTCTGAGTCCA-----GCACGACCTCCAGTGGGCGCAG 1131
QY 150 InGlnHisProAlaValGlnIYsProThrProProValValValIYsYsProThrP 170
```

```
DB 1132 CACAGCCACCAACTCTGAGTCCAGCAGCACTCTCCAGTGGGCTAGCAGACCACTC 1191
QY 170 roThrProProValValInGlnPro--AlaProValAlaProProVal-----T 186
DB 1192 TGACTCCA-----GCACAACTCTCCAGTGGGCGCGCAGCAGCAGCACTCTGAGTCCAG 1245
QY 186 hGlnAlaProPheAlaThrGlySerSergIYValMetGlnPheArgTrgProValGlyA 206
DB 1246 CACAGTGTCCAGTGGGATCCAGCAGACAGTCCCAATTTGAGTCCAGCAGCACTCTC-AGTG 1304
QY 206 lArThAsnProValValArgArgPheGlyThrAlaThrValAlaGlySerThrValThrs 226
DB 1305 GGGCCAAAC-----ACAGCCACCACTCTGAGTCCAGTCCAGTCCAGTCCAGTCC 1343
QY 226 eRAsnGlyMetTrpPheSergIYArgAspGlyAspLeuIleAsnAlaSerAsnAlaGlyT 246
DB 1344 CCAGTGGG-----GCCAACACAGCCA 1364
QY 246 hVAlIleGlnAlaAspHisAsnMetAspGlyAlaSerIleValIleGlnHis----- 263
DB 1365 CCACTCTGAGTCCAGCAGCAGTCCAGTGGGCGCAGCAGTCCAGCAGCACTCTGAGTCCA 1424
QY 264 -----ThrAsnGlyPheValSerSergIYrIleHisIleIYsAspAlaGlnValYsThrg 282
DB 1425 GCACAACTCTCCAGTGGGCTCAGCAGCCACCACTCTGAGTCCAGCAGCACTCTGAGTCCAGTGG 1484
QY 282 lYAspThrValArgThrGlyGlnArgIleAlaSerMetIYsAsnGlnProSergIYAla 302
DB 1485 GGAGTTCAGCAGCCACCACTCTGAGTCCAGCAGCACTCTCCAGTGGGCGCAGCAGCCA 1544
QY 302 lAlaPheGlnPheArgIleSerArgAsnGlyVal 313
DB 1545 CCAACTCTGAGTCTAGCAGACAGTTCAGTGGGATC 1579

RESULT 26
US-10-063-616-99
; Sequence 99, Application US/10063616
; Publication No. US20030013855A1
; GENERAL INFORMATION:
; APPLICANT: Eaton, Dan L.
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3230P1C1
; CURRENT APPLICATION NUMBER: US/10/063,616
; CURRENT FILING DATE: 2002-05-03
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 170
; SEQ ID NO 99
; LENGTH: 2436
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-063-616-99

Alignment Scores:
Pred. No.: 0.00289 Length: 2436
Score: 122.50 Matches: 78
Percent Similarity: 39.04% Conservative: 52
Best Local Similarity: 23.42% Mismatches: 145
Query Match: 7.34% Indels: 60
DB: Gaps: 12

US-10-018-706-2 (1-322) x US-10-063-616-99 (1-2436)
QY 6 AlaIleAsnSerGlnAsnGlnIysProIleuYsArGleuGlyLeuIlePheGlyValIle 25
```

```

Db      686 GCCACCAACTGTGAGTCCAGACAGTGTCCAGTAGGGCCAGCACTGCCAACCTGTAG 745
Qy      26 ThrThrCysIleLeuAlaGlyCysAlaSerIleProThrTyrAsnSerThrSerGlySer 45
Db      746 TCTAGCACACTCTCCAGTGGGGCC-----AGACAGCCACCACTCT 787
Qy      46 GlySerHisArgThrSerGlySerGlyLeuAlaIleGlySerGlnValIleThrAsp 65
Db      788 GACTCCAGACACACTCCAGTGGGGCTAGACAGCCACCACTGAGTCCAGACCAACC 847
Qy      66 SerGlnGlyValProAsnArgTyrGln-----VallyGlnGlyAsp 79
Db      848 TCCAGTGGGGCCAG-CACAGCCACCACTGAGTCCAGACAGTGTCCAGTAGGGCCAG 906
Qy      80 ThrValSerIleIleAlaGlnArgTyrGlyLeuAsnTyrPargGluIleGlyHis----- 97
Db      907 CACTGCCCAACTGTGAGTCCAGACCACTCCAGTGGGGCCAGACACAGCCACCACTC 966
Qy      98 -----IleAsnAsnLeu-AsnSerSerTyrThrIleTyrThrGlyGlnTyrPleuThrIle 115
Db      967 TGAAGTCAGAAAGACCTCCAAATGGGGCTGGGACACAGCCACCACTGAGTCCAGACGAC 1026
Qy      115 UTPrSerGlyAspLeuIleValArg-----GluArgSerIleSerSerGlyValAs 132
Db      1027 CTCCTGAGGGGCCAGACAGCCACCACTGAGTCCAGTCCAGTGTCCAGTGGGGCCAG 1086
Qy      132 nThrAla-HisThrProSerProValAlaValGlnSerSerArgProProVal-----G 150
Db      1087 CACTGCCCAACTGTGAGTCCA-----GCCAGACTCCAGTGGGGCCAG 1131
Qy      150 IngInHisProAlaValGlnIleProThrProProValValValIleValIleProThrP 170
Db      1132 CACAGCCACCAACTGTGAGTCCAGACAGACACCTCCAGTGGGGCTAGACAGCCACCACTC 1191
Qy      170 roThrProProValValGlnIlePro--AlaProValAlaProProVal-----T 186
Db      1192 TGACTTCCA-----GCACCACTCCAGTGGGGCCAGCCAGCCACCACTGTGAGTCCAG 1245
Qy      186 hrGlnAlaProPheAlaThrGlySerSerGlyValMetGlnPheArgTyrProValGlyA 206
Db      1246 CACAGTGTCCAGTGGGATGACAGACAGTCAACATCTGAGTCCAGACCACTCC-AGTG 1304
Qy      206 IarHisnProValValArgArgPheGlyThrAlaThrValAlaGlySerThrValIleHis 226
Db      1305 GGGGCCAAC-----ACAGCCACCACTGTGAGTCCAGTACGACT 1343
Qy      226 eraHnGlyMetTyrPheSerGlyArgAspGlyAspLeuIleAsnAlaSerAsnAlaGlyT 246
Db      1344 CCAGTGGG-----GCCAACACAGCCA 1364
Qy      246 hrValIleGlnAlaAspHisAsnMetArgValAlaSerIleValIleGlnHis----- 263
Db      1365 CCAACTGTGAGTCCAGACAGTGTCCAGTGGGGCCAGACCTGCCAACCACTGTGAGTCCA 1424
Qy      264 -----ThrAsnGlyPheValSerSerTyrIleHisIleGlyAspAlaGlnValIleThrG 282
Db      1425 GCACCAACTCCAGTGGGGCTAGACAGCCACCACTGAGTCCAGACCACTCCAGTCCAGT 1484
Qy      282 IAspThrValArgThrGlyGlnArgIleAlaSerMetIleAsnGlnProSerGlyAlaA 302
Db      1485 GGGGTAGACAGACGCCCAACTGTGAGTCCAGACCACTCCAGTGGGGCCAGACAGCCA 1544
Qy      302 IaleuPheGluPheArgIleSerArgAsnGlyVal 313
Db      1545 CCAACTGTGAGTCCAGACAGTGTCCAGTGGGATC 1579

```

```

; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C50
; CURRENT APPLICATION NUMBER: US/10/175,737
; PRIORITY FILING DATE: 2002-06-19
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 309
; LENGTH: 2436
; TYPE: DNA
; ORGANISM: Homo Sapien
; US-10-175-737-309

Alignment Scores:
Pred. No.: 0.00289 Length: 2436
Score: 122.50 Matches: 78
Percent Similarity: 39.04% Conservative: 52
Best Local Similarity: 23.42% Mismatches: 145
Query Match: 7.34% Indels: 60
DB: Gaps: 12

US-10-018-706-2 (1-322) x US-10-175-737-309 (1-2436)
Qy      6 AlaIleAsnSerGlnAsnGlnIleProIleAspLeuGlyLeuIlePheGlyValIle 25
Db      686 GCCACCAACTGTGAGTCCAGACAGTGTCCAGTGGGGCCAGCACTGCCAACCTGTAG 745
Qy      26 ThrThrCysIleLeuAlaGlyCysAlaSerIleProThrTyrAsnSerThrSerGlySer 45
Db      746 TCTAGCACACTCTCCAGTGGGGCC-----AGACAGCCACCACTCT 787
Qy      46 GlySerHisArgThrSerGlySerGlyLeuAlaIleGlySerGlnValIleThrAsp 65
Db      788 GACTCCAGACACACTCCAGTGGGGCTAGACAGCCACCACTGAGTCCAGACCAACC 847
Qy      66 SerGlnGlyValProAsnArgTyrGln-----VallyGlnGlyAsp 79
Db      848 TCCAGTGGGGCCAG-CACAGCCACCACTGAGTCCAGACAGTGTCCAGTAGGGCCAG 906
Qy      80 ThrValSerIleIleAlaGlnArgTyrGlyLeuAsnTyrPargGluIleGlyHis----- 97
Db      907 CACTGCCCAACTGTGAGTCCAGACCACTCCAGTGGGGCCAGACACAGCCACCACTC 966
Qy      98 -----IleAsnAsnLeu-AsnSerSerTyrThrIleTyrThrGlyGlnTyrPleuThrIle 115
Db      967 TGAAGTCAGAAAGACCTCCAAATGGGGCTGGGACACAGCCACCACTGAGTCCAGACGAC 1026
Qy      115 UTPrSerGlyAspLeuIleValArg-----GluArgSerIleSerSerGlyValAs 132
Db      1027 CTCCTGAGGGGCCAGACAGCCACCACTGAGTCCAGTCCAGTGTCCAGTGGGGCCAG 1086
Qy      132 nThrAla-HisThrProSerProValAlaValGlnSerSerArgProProVal-----G 150
Db      1087 CACTGCCCAACTGTGAGTCCA-----GCCAGACTCCAGTGGGGCCAG 1131
Qy      150 IngInHisProAlaValGlnIleProThrProProValValValIleValIleProThrP 170
Db      1132 CACAGCCACCAACTGTGAGTCCAGACAGACACCTCCAGTGGGGCTAGACAGCCACCACTC 1191
Qy      170 roThrProProValValGlnIlePro--AlaProValAlaProProVal-----T 186
Db      1192 TGACTTCCA-----GCACCACTCCAGTGGGGCCAGCCAGCCACCACTGTGAGTCCAG 1245
Qy      186 hrGlnAlaProPheAlaThrGlySerSerGlyValMetGlnPheArgTyrProValGlyA 206

```

RESULT 27
 US-10-175-737-309
 ; Sequence 309, Application US/10175737
 ; Publication No. US20030013153A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Baker, Kevin P.
 ; APPLICANT: Chen, Jian

```

Db      1246 CACAGTCTCAGTGGAGATCAGCAGCAGTCCAAATTCAGTCCAGCACACCTCC-AGTG 1304
      206 lATHAsnProvalValAlArGArpHeGlyThrAlaThValAlaGlySerThrValThrs 226
      1305 GGGCCAAAC-----ACAGCCACCACTCTGAGTCCAGTCACTGACT 1343
Qy      226 eYAsnGlyWetrPrPheSerGlyArGAspGlyAspLeuIlEAsnAlaSerSnaIaGlyT 246
      1344 CCAGTGGG-----GCCAACACAGCCA 1364
Qy      246 hVAlIleGlnAlaAspHisAsnMetAspGlyAlaSerIleValIleGlnHis----- 263
      1365 CCAACTGTGAGTCCAGCAGCAGTGTCCAGTGGGCCAGCAGCAGCCCAACTGTGATCCA 1424
Qy      264 -----ThrAsnGlyPheValIleSerSerTyrlIeHisIleLeuAspAlaGlnValIleThrs 282
      1425 GCACAACTCTCAGTGGAGTGGGAGTCCAGCAGCAGCAGCAGCAGCAGCAGCAGTCCAGT 1484
Qy      282 lYAspThrValAlArGThrGlyGlnAlArGleAlaSerMetLysAsnGlnProSerGlyAla 302
      1485 GGGCTAGCAGCAGCCCAACTCTGTGACTCCAGCAGCAACTCTCCAGTGGCCAGCAGCCA 1544
Qy      302 lAlaLeuPheGlnPheArGlyIleSerArGAsnGlyVal 313
      1545 CCAACTGTGAGTCTAGCAGCAGTGTCCAGTGGGATC 1579

```

RESULT 28
US-10-063-502-99
Sequence 99, Application US/10063502
Publication No. US20030023042A1

GENERAL INFORMATION:
APPLICANT: Eaton, Dan L.
APPLICANT: Filvaroff, Ellen
APPLICANT: Gerlitsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.

TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3230R1C1
CURRENT FILING DATE: 2002-05-01
CURRENT APPLICATION NUMBER: US/10/063,502
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 170
SEQ ID NO 99
LENGTH: 2436
TYPE: DNA
ORGANISM: Homo Sapien
US-10-063-502-99

Alignment Scores:
Pcted. No.: 0.00289
Score: 122.50
Percent Similarity: 39.04%
Best Local Similarity: 23.42%
Query Match: 7.34%
DB: 9
Gaps: 12

US-10-018-706-2 (1-322) x US-10-063-502-99 (1-2436)

```

Qy      6 AlAlIAsnSerGlnAsnGlnLysProlIeYAsrLeuGlyLeuIllePheGlyValIle 25
      686 GCCACCAACTCTGAGTCCAGCAGCAGTGTCCAGTGGGCCAGCAGCAGCAGCAGCAGTCCAG 745
Qy      26 ThrThrCyAlleuAlaGlyCysAlaSerLysProThrTyAsnSerThrsSerGlySer 45
      746 TCTAGCAGCACTCTCCAGTGGGCC-----AGCAGCCCACTCAACTCT 787
Qy      46 GlySerHisArGThrSerGlySerGlyGlyLeuAlaIleGlySerGlnValIleThrs 65

```

```

Db      788 GACTCCAGCAGCACTCTCAGTGGGCTAGCAGCAGCAGCAGCAGCAACTCTGAGTCCAGCAGCAC 847
      66 SerGlnGlyValProAsnArGlyGln-----ValIleGlnGlyAsp 79
      848 TCCAGTGGGCCAG-CACAGCCACCACTCTGAGTCCAGCAGCAGTGTCCAGTGGGCCAG 906
Qy      80 ThrValSerIleAlaGlnArGlyGlyLeuSnrTPArGlyIleGlyHis----- 97
      907 CACTGCCCAACTCTGAGTCCAGCAGCAACTCTCCAGTGGGCCAGCAGCAGCAGCAGCAGTCC 966
Qy      98 -----lAsnAsnLeu-AsnSerSerTyrlleThrGlyGlnTrpLeuThle 115
      967 TGAAGTCCAGACAGACTCTCCAAATGGGCTGGCAGCAGCAGCAGCAACTGTAGTCCAGCAGAC 1026
Qy      115 utrPserGlyAspLeuLysValArg-----GlnArGserIleSerSerGlyValAs 132
      1027 CTCAGTGGGCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGTGTCCAGTGGGCCAG 1086
Qy      132 nThrAla-HisThrProSerProValAlaValGlnSerSerArGProProVal-----G 150
      1087 CACTGCCCAACTCTGAGTCCA-----GCAGACTCTCAAGTGGGCCAG 1131
Qy      150 lngIleHisProAlaValGlnLysProThrProProValValAlaValLysLysProThrP 170
      1132 CACAGCCCAACTCTGAGTCCAGCAGCAGCAGCAGTGTGGGCTAGCAGCAGCAGCAGCAGTCC 1191
Qy      170 rOThrProProValValGlnGlnPro--AlaProValAlaProProVal-----T 186
      1192 TGACTCCA-----GCACAACTCTCAGTGGGCCGCGCAGCAGCAGCAACTCTGAGTCCAG 1245
Qy      186 hrgIuAlaProPheAlaThrGlySerSerGlyValMetGlnPheArGlyTrProValGly 206
      1246 CACAGTGTCCAGTGGAGTCCAGCAGCAGTCCCAATCTGAGTCCAGCAGCAGCAGTCC-AGTG 1304
Qy      206 lATHAsnProvalValAlArGArpHeGlyThrAlaThValAlaGlySerThrValThrs 226
      1305 GGGCCAAAC-----ACAGCCACCACTCTGAGTCCAGTCACTGAGTCCAGTCCAGTCC 1343
Qy      226 eYAsnGlyWetrPrPheSerGlyArGAspGlyAspLeuIlEAsnAlaSerSnaIaGlyT 246
      1344 CCAGTGG-----GCCAACACAGCCA 1364
Qy      246 hVAlIleGlnAlaAspHisAsnMetAspGlyAlaSerIleValIleGlnHis----- 263
      1365 CCAACTGTGAGTCCAGCAGCAGTGTCCAGTGGGCCAGCAGCAGCAGCAGCAGTCCAGTCCA 1424
Qy      264 -----ThrAsnGlyPheValIleSerSerTyrlIeHisIleLeuAspAlaGlnValIleThrs 282
      1425 GCACAACTCTCAGTGGAGTGGGAGTCCAGCAGCAGCAGCAGCAGCAGCAGCAGTCCAGT 1484
Qy      282 lYAspThrValAlArGThrGlyGlnAlArGleAlaSerMetLysAsnGlnProSerGlyAla 302
      1485 GGGCTAGCAGCAGCCCAACTCTGTGACTCCAGCAGCAACTCTCCAGTGGCCAGCAGCCA 1544
Qy      302 lAlaLeuPheGlnPheArGlyIleSerArGAsnGlyVal 313
      1545 CCAACTGTGAGTCTAGCAGCAGTGTCCAGTGGGATC 1579

```

RESULT 29
US-10-173-706-309
Sequence 309, Application US/10173706
Publication No. US20030022293A1

GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jian
APPLICANT: Desnoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Colin K.

```

; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C7
; CURRENT APPLICATION NUMBER: US/10/173,706
; PRIOR FILING DATE: 2002-06-17
; PRIOR APPLICATION REMOVED - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 309
; LENGTH: 2436
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-173-706-309

Alignment Scores:
Pred. No.: 0.00289 Length: 2436
Score: 122.50 Matches: 78
Percent Similarity: 39.04% Conservative: 52
Best Local Similarity: 23.42% Mismatches: 145
Query Match: 7.34% Indels: 60
Gaps: 12

US-10-018-706-2 (1-322) x US-10-173-706-309 (1-2436)
QY 6 AAlaIleAnSerGlnAnGlnLysProIleLysArgLeuGlyLeuIlePheGlyValIle 25
Db 686 GCCACCAACTCTGAGTCCAGACAGTCTCCAGTGGGCGACACCTGCCACCACTCTGAG 745
QY 26 ThrThrCysIleLeuAlaGlyCysAlaSerLysProThrTyrrAnSerThrSerGlySer 45
Db 746 TCTAGCACACTCTCCAGTGGGGCC-----AGCACGCCACCACTCT 787
QY 46 GlySerHisArgThrSerGlySerGlyLeuAlaIleGlySerGlnValIleThrAsp 65
Db 788 GACTCCAGCACCACTCTCCAGTGGGGCTAGACAGCCCACTGAGTCCAGCACCACTC 847
QY 66 SerGlnGlyValProAnArgTyrGln-----ValIleGlnGlyAsp 79
Db 848 TCCAGTGGGGCGCAG-CACAGCCACCACTCTGAGTCCAGCACAGTCTCACTAGGGCCAG 906
QY 80 ThrValSerLysIleAlaGlnArgTyrGlyLeuAnThrPargIleGlyHis----- 97
Db 907 CACTGCCACCACTCTGAGTCCAGACCACTCCAGTGGGGCGACAGCACGCCCACTC 966
QY 98 -----IleAnSerLysLeuAnSerSerTyrThrIleTyrThrGlyGlnTrpLeuThrIle 115
Db 967 TGAGTCCAGAACGACTCTCAATGGGGCTGACAGCCCACTCTGAGTCCAGACGAC 1026
QY 115 uTrpSerGlyAspLeuLysValArg-----GlnArgSerIleSerSerGlyValAsp 132
Db 1027 CTCGAGTGGGGCGCAGCACAGCCCACTGAGTCTCCAGCACAGTCTCACTAGGGCCAG 1086
QY 132 nThrAla-HisThrProSerProValAlaValGlnSerSerArgProProVal-----G 150
Db 1087 CACTGCCACCACTCTGAGTCCA-----GCCAGACTCTCACTAGGGGGCCAG 1111
QY 150 lnglnHisProAlaValGlnLysProThrProProValValValLysLysProThrP 170
Db 1132 CACAGCCACCACTCTGAGTCCAGACGACTCTCCAGTGGGGCTGACAGCACGCCCACTC 1191
QY 170 roThrProProValValGlnGlnPro---AlaProValAlaLysProProVal-----T 186
Db 1192 TGACTTCCA-----GCACAACTCTCACTAGGGGGCGGCGCACAGCCCACTCTGAGTCCAG 1245
QY 186 hrGlnAlaProPheAlaThrGlySerSerGlyValMetGlnPheArgTyrProValGly 206
Db 1246 CACAGTCCAGTGGGATCCAGACAGTCCACCACTCTGAGTCCAGACCACTCTC-AGTG 1304
QY 206 lathrAnProValValArgArgPheGlyThrAlaThrValAlaGlySerThrValThrS 226
Db 1305 GGGGCCAAC-----ACAGCCACCACTCTGAGTCCAGTCCAGTCCAGTCCAGTCC 1343

```

```

QY 226 eRAnGlyMetTrpPheSerGlyArgAspGlyAspLeuIleAsnAlaSerAsnAlaGlyT 246
Db 1344 CCACTGGG-----GCCAACAGACCA 1364
QY 246 hrValIleGlnAlaAspHisMetAspGlyAlaSerIleValIleGlnHis----- 263
Db 1365 CCAACTGAGTCCAGACAGAGTCTCCAGTGGGGCGGCGCACACTCTGAGTCCAGTCC 1424
QY 264 -----ThrAnGlyPheValSerSerTyrIleHisIleLysAspAlaGlnValLysThrG 282
Db 1425 GCACAACTCTCCAGTGGGGCTGACAGCACGACCACTCTGAGTCCAGACCACTCTCAGTG 1484
QY 282 lYAspThrValArgThrGlyGlnArgIleAlaSerMetLysAsnGlnProSerGlyVal 302
Db 1485 GGGGTACACAGACGCCCACTCTGAGTCCAGACCACTCTCAGTCCAGTCCAGTCCAGTCC 1544
QY 302 lalauPheGlyPheArgIleSerArgAsnGlyVal 313
Db 1545 CCAACTCTGAGTCTAGACAGTCTCCAGTGGGATC 1579

RESULT 30
US-10-175-738-309
; Sequence 309, Application US/10175738
; Publication No. US20030022294A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C45
; CURRENT APPLICATION NUMBER: US/10/175,738
; PRIOR FILING DATE: 2002-06-19
; PRIOR APPLICATION REMOVED - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 309
; LENGTH: 2436
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-175-738-309

Alignment Scores:
Pred. No.: 0.00289 Length: 2436
Score: 122.50 Matches: 78
Percent Similarity: 39.04% Conservative: 52
Best Local Similarity: 23.42% Mismatches: 145
Query Match: 7.34% Indels: 60
Gaps: 12

US-10-018-706-2 (1-322) x US-10-175-738-309 (1-2436)
QY 6 AAlaIleAnSerGlnAnGlnLysProIleLysArgLeuGlyLeuIlePheGlyValIle 25
Db 686 GCCACCAACTCTGAGTCCAGACAGTCTCCAGTGGGCGACACCTGCCACCACTCTGAG 745
QY 26 ThrThrCysIleLeuAlaGlyCysAlaSerLysProThrTyrrAnSerThrSerGlySer 45
Db 746 TCTAGCACACTCTCCAGTGGGGCC-----AGCACGCCACCACTCT 787
QY 46 GlySerHisArgThrSerGlySerGlyLeuAlaIleGlySerGlnValIleThrAsp 65
Db 788 GACTCCAGCACCACTCTCACTAGTGGGGCTAGACAGCCCACTCTGAGTCCAGCACCACTC 847
QY 66 SerGlnGlyValProAnArgTyrGln-----ValIleGlnGlyAsp 79

```

Db 848 TCACAGTGGGGCCAG-CACAGCCACCACTCTGATCTCAGACAGACAGTGTCCAGTAGGGCCAG 906
QY 80 ThrValSerLysIleAlaGlnArgTyrGlyLeuAenTPRArgGluIleGlyHis----- 97
Db 907 CACTGCCCACTCTGAGTCCAGACCACTCCAGTGGGGCCAGACAGCCACCACTCTC 966
QY 98 -----IleAsnAsnLeu-AsnSerSerTyrThrIleTyrThrGlyGlnTyrPhe 115
Db 967 TGAGTCCAGACCACTCTGAGTCCAGACCACTCTGAGTCCAGACCACTCTCAGTCCAGAC 1026
QY 115 uTPSerGlyAspLeuLysValArg-----GluArgSerIleSerSerGlyValAs 132
Db 1027 CTCAGTGGGGCCAGACAGCCACCACTCTGATCTCAGACAGTGTCCAGTGGGGCCAG 1086
QY 132 nThrAla-HisThrProSerProValAlaValAlaGlnSerSerArgProProVal-----G 150
Db 1087 CACTGCCCACTCTGAGTCCAGACCACTCTGAGTCCAGACCACTCTCAGTGGGGCCAG 1131
QY 150 IngIHisProAlaValAlaGlnLysProThrProProValAlaValAlaLysLysProThrP 170
Db 1132 CACAGCCACCACTCTGAGTCCAGACCACTCTCAGTGGGGCCAGACAGCCACCACTCTC 1191
QY 170 roThrProProValValGlnGlnPro--AlaProValAlaProProVal-----T 186
Db 1192 TGACTCCA-----GCACACCTCTCAGTGGGGCCAGACAGCCACCACTCTGAGTCCAG 1245
QY 186 hrcGlnAlaProPheAlaThrGlySerSerGlyValMetGlnPheArgTyrProValGlyA 206
Db 1246 CACAGTGTCCAGTGGGATCCAGACAGTCCCAATTCGATCCAGACCACTCTCAGTCCAG 1304
QY 206 lATHrAnProValValArgArgPheGlyThrAlaThrValAlaGlySerThrValThrs 226
Db 1305 GGGCCCAAC-----ACAGCCACCACTCTGAGTCCAGTCCAGTCCAGTCCAGTCCAG 1343
QY 226 eArgnGlyMetTrpPheSerGlyArgAspGlyAspLeuIleAsnAlaSerAsnAlaGlyT 246
Db 1344 CCAAGTGGG-----GCCAACACAGCCA 1364
QY 246 hrValIleGlnAlaAspHisAsnMetAspGlyAlaSerIleValIleGlnHis----- 263
Db 1365 CCAACTCTGATCTCAGACAGTGTCCAGTGGGGCCAGCACTGCCACCACTCTGAGTCCA 1424
QY 264 -----ThrAngnGlyPheAlaSerSerTyrIleHisIleLysAspAlaGlnValLysThng 282
Db 1425 GCACAACTCTCAGTGGGATCCAGACAGCCACCACTCTGAGTCCAGACCACTCTCAGTCCAG 1484
QY 282 lYAspThrValArgThrGlyGlnArgIleAlaSerMetLysAsnGlnProSerGlyValAs 302
Db 1485 GGGCTGACACAGCCACCACTCTGAGTCCAGACCACTCTCAGTGGGGCCAGACAGCCA 1544
QY 302 lAlaLeuPheGluPheArgIleSerArgAsnGlyVal 313
Db 1545 CCAACTCTGATCTCAGACAGTGTCCAGTGGGATC 1579

RESULT 31
US-10-175-752-309
Sequence 309, Application US/10175752
Publication No. US2003022295A1
GENERAL INFORMATION:

APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jian
APPLICANT: Desnoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Matanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3430R1C60

US-10-018-706-2 (1-322) x US-10-175-752-309 (1-2436)
QY 6 AlaIleAsnSerGlnAsnGlnLysProIleLysArgLeuGlyLeuIlePheGlyValIle 25
Db 686 GCCACCACTCTGAGTCCAGACCACTCTGAGTCCAGACCACTCTGAGTCCAGACCACTCTGAG 745
QY 26 ThrThrCysIleLeuAlaGlyCysAlaSerLysProThrTyrAsnSerThrSerGlySer 45
Db 746 TCTAGCACACTCTCCAGTGGGGC-----AGCACAGCCACCACTCT 787
QY 46 GlySerHisArgThrSerGlySerGlyValLeuAlaIleGlySerGlnValIleThrAsp 65
Db 788 GACTCCAGACCACTCTCAGTGGGCTGACAGACAGCCACCACTCTGAGTCCAGACCACTCTC 847
QY 66 SerGlnGlyValProAsnArgTyrGln-----ValGlnGlnLysAsp 79
Db 848 TCCAGTGGGGCCAG-CACAGCCACCACTCTGAGTCCAGACCACTCTCAGTGGGGCCAG 906
QY 80 ThrValSerLysIleAlaGlnArgTyrGlyLeuAenTPRArgGluIleGlyHis----- 97
Db 907 CACTGCCCACTCTGAGTCCAGACCACTCTCAGTGGGGCCAGACAGCCACCACTCTCAGTCCAG 966
QY 98 -----IleAsnAsnLeu-AsnSerSerTyrThrIleTyrThrGlyGlnTyrPhe 115
Db 967 TGAGTCCAGACCACTCTCAGTGGGGCCAGACAGCCACCACTCTGAGTCCAGACCACTCTC 1026
QY 115 uTPSerGlyAspLeuLysValArg-----GluArgSerIleSerSerGlyValAs 132
Db 1027 CTCAGTGGGGCCAGACAGCCACCACTCTGATCTCAGACAGTGTCCAGTGGGGCCAG 1086
QY 132 nThrAla-HisThrProSerProValAlaValAlaGlnSerSerArgProProVal-----G 150
Db 1087 CACTGCCCACTCTGAGTCCAGACCACTCTGAGTCCAGACCACTCTCAGTGGGGCCAG 1131
QY 150 IngIHisProAlaValAlaGlnLysProThrProProValAlaValAlaLysLysProThrP 170
Db 1132 CACAGCCACCACTCTGAGTCCAGACCACTCTCAGTGGGGCCAGACAGCCACCACTCTC 1191
QY 170 roThrProProValValGlnGlnPro--AlaProValAlaProProVal-----T 186
Db 1192 TGACTCCA-----GCACACCTCTCAGTGGGGCCAGACAGCCACCACTCTGAGTCCAG 1245
QY 186 hrcGlnAlaProPheAlaThrGlySerSerGlyValMetGlnPheArgTyrProValGlyA 206
Db 1246 CACAGTGTCCAGTGGGATCCAGACCACTCTGATCCAGACCACTCTCAGTCCAGACCACTCTC 1304
QY 206 lATHrAnProValValArgArgPheGlyThrAlaThrValAlaGlySerThrValThrs 226
Db 1305 GGGCCCAAC-----ACAGCCACCACTCTGAGTCCAGTCCAGTCCAGTCCAGTCCAG 1343
QY 226 eArgnGlyMetTrpPheSerGlyArgAspGlyAspLeuIleAsnAlaSerAsnAlaGlyT 246
Db 1344 CCAAGTGG-----GCCAACACAGCCA 1364
QY 246 hrValIleGlnAlaAspHisAsnMetAspGlyAlaSerIleValIleGlnHis----- 263

CURRENT APPLICATION NUMBER: US/10175,752
CURRENT FILING DATE: 2002-06-19
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 612
SEQ ID NO 309
LENGTH: 2436
TYPE: DNA
ORGANISM: Homo Sapien
US-10-175-752-309

Alignment Scores:
Pred. No.: 0.00289 Length: 2436
Score: 122.50 Matches: 78
Percent Similarity: 39.04% Conservative: 52
Best Local Similarity: 23.42% Mismatches: 145
Query Match: 7.34% Indels: 60
Gaps: 12


```

Db      1365 CCAACTCTGAGTCCAGACACAGTGTCAGGTGGGGCCAGACACTGCCAACCTCTGAGTCCA 1422
Oy      264 -----ThraSnGLyPheValSerSeryTrIleHisILeYsaPaLaGIInVallysrhg 282
          |||:::|||||||:::|||||:::
Db      1425 GCACAACCTCAGTGGGGGTGAGCACAGGCACCACAACTGTGATGCCAGACCAACTCGAAGTg 1488
Oy      282 lYaSPThVaLArghrGelYglInAgllLeAlaSeRMeCLyaSnGLInProSerGlYalaa 302
          |||:::|||||:::|||||:::
Db      1485 GGGCTGACAGACGCCCAACTCTGTACTCCAGACCAACTCCAGTAGAGGGCCAGACAGCCA 1544
Oy      302 lalaupheGLuPheArgrlIeserArgaSnGLYval 313
          |||:::|||||:::
Db      1545 CCAACTCTGAGTCTAGCACACAGTGTCCAGTTGGGATC 1579

RESULT 32
US-10-176-482-309
Sequence 309, Application US/10176482
Publication No. US2003002296A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jian
APPLICANT: Desnoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Aueclin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: ACIDS ENCODING THE SAME
CURRENT APPLICATION NUMBER: US/10/176, 482
CURRENT FILING DATE: 2002-06-20
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 612
SEQ ID NO 309
LENGTH: 2436
TYPE: DNA
ORGANISM: Homo Sapien
US-10-176-482-309

Alignment Scores:
Pred. No.: 0.00289 Length: 2436
Score: 122.50 Matches: 78
Percent Similarity: 39.04% Conservative: 52
Best Local Similarity: 23.42% Mismatches: 145
Query Match: 7.34% Indels: 60
DB: Gaps: 12

US-10-018-706-2 (1-322) x US-10-176-482-309 (1-2436)
Oy      6 AlaIleAsnserGlnAsnGLInLyBProIlleYsaGLeuGLYleuIllePhelYValIle 25
          |||:::|||||||:::|||||:::
Db      686 GCCACCAACTGTGAgtCCAGACAGACTGTCCAGTAAGGGCCAGACAGCACTGAG 745
Oy      26 ThrThrCySlleLeuAlaGLYCysAlaSerLYSProThrTYAsSerThrserylYser 45
          |||:::|||||:::|||||:::
Db      746 TCTACACACACTCTCCAGTGGGCC-----AGCAGGCCACCAACTCT 787
Oy      46 GlySerHIsaRGThrserylSerylGlyLeuAlaIleGlySerGLInValIIeThrasP 65
          |||:::|||||:::|||||:::
Db      788 GACTCCAGACCAACTCCAGTGGGGCTAGCACAGGCACCACAACTGTGATGCCAGCACACACC 847
Oy      66 SerGLInGLYalProASnaRGTYrlN-----VallySGInGLYasP 79
          |||:::|||||:::|||||:::
Db      848 TCCAGTGGGGCCAG-CACAGCCACCAACTGTGATGCCAGACAGACTGCCAGTAGGGCCAG 906
Oy      80 ThraValseRylleAlaGLInArGrTyrlYleuAsnTPraGInLIleGLYHis----- 97
          |||:::|||||:::|||||:::
Db      907 CACTGCCACCAACTGTGATGCCAGACCAACTCCAGTGGGGCCAGACAGCCACCAACTC 966

```

LENGTH: 2436
TYPE: DNA
ORGANISM: Homo Sapien
US-10-176-757-309

Alignment Scores:
Pred. No.: 0.00289 Length: 2436
Score: 122.50 Matches: 78
Percent Similarity: 39.04% Conservative: 52
Best Local Similarity: 23.42% Mismatches: 145
Query Match: 7.34% Indels: 60
Gaps: 12

US-10-018-706-2 (1-322) x US-10-176-757-309 (1-2436)

```

QY 6 AAlaIleAsnSerGlnAsnGlnLysProIleYsArgLeuGlyLeuIlePheGlyValIle 25
DB 686 GCACCACTCTGAGTCCAGCAGCAGTCTGAGTCCAGTCCAGCAGCAGCAGCAGCAGCAG 745
QY 26 ThrThrCysIleLeuAlaGlyCysAlaSerLysProThrTyraSerThrSerGlySer 45
DB 746 TCTAGCAGCAGTCTCCAGTGGGGCC-----AGCAGCAGCAGCAGCAGCAGCAGCAGC 787
QY 46 GlySerHisArgThrSerGlySerGlyLeuAlaIleGlySerGlnValIleThrAsp 65
DB 788 GACTCCAGCAGCAGCAGCAGTCCAGTGGGGCTGACAGCAGCAGCAGCAGCAGCAGCAGC 847
QY 66 SerGlnGlyValProAsnArgTyrgln-----ValIleGlnGlyAsp 79
DB 848 TCCAGTGGGGCCAG-CACAGCCAGCAGCAGTCTGAGTCCAGCAGCAGTCCAGTGGGGCCAG 906
QY 80 ThrValSerLysIleAlaGlnArgTyrglyLeuAsnTrpArgGluIleGlyHis----- 97
DB 907 CACTGCAGCAGCAGTCTGAGTCCAGCAGCAGCAGTCTGAGTGGGGCCAGCAGCAGCAGC 966
QY 98 -----IleAsnAsnLeuAsnSerSerTyThrIleTyThrGlyGlnTrpLeuThrIle 115
DB 967 TGAAGTCCAGAGCAGTCTCCAGTGGGGCTGAGCAGCAGCAGCAGCAGTCTGAGTCCAGCAG 1026
QY 115 UTPrSerGlyAspLeuLysValArg-----GluArgSerIleSerSerGlyValAs 132
DB 1027 CTCACAGTGGGGCCAGCAGCAGCAGCAGCAGTCTGAGTCCAGCAGTCTGAGTGGGGCCAG 1066
QY 132 nThrAla-HisThrProSerProValAlaValGlnSerSerTrpProVal-----G 150
DB 1087 CACTGCAGCAGCAGTCTGAGTCC-----GCAGCAGTCTCCAGTGGGGCCAG 1131
QY 150 IlnIlnIleProAlaValAlaGlnLysProThrProProValValValLysLysProThr 170
DB 1132 CACAGCAGCAGCAGTCTGAGTCCAGCAGCAGTCTGAGTGGGGCTGAGCAGCAGCAGCAG 1191
QY 170 roThrProProValValGlnGlnPro--AlaProValAlaProProVal-----T 186
DB 1192 TGAATCCA-----GCAGCAGTCTCCAGTGGGGCCAGCAGCAGCAGCAGTCTGAGTCCAG 1245
QY 186 hrcIuAlaProPheAlaThrGlySerSerGlyValMetGlnPheArgTyProValGlyA 206
DB 1246 CAGAGTCTCCAGTGGATCCAGCAGTCCAGCAGTCTGAGTCCAGCAGCAGCAGCAGTCT 1304
QY 206 lAThrAsnProValValArgArgPheGlyThrAlaThrValAlaGlySerThrValThr 226
DB 1305 GGGCCAGC-----ACAGCCAGCAGCAGTCTGAGTCCAGTCCAGTCCAGTCCAGT 1343
QY 226 eArgnGlyMetTrpPheSerGlyArgAspGlyAspLeuIleAsnAlaSerAsnAlaGlyT 246
DB 1344 CCAAGTGG-----GCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1364
QY 246 hrcAlaIleGlnAlaAsnHisAsnMetAspGlyAlaSerIleValIleGlnHis----- 263
DB 1365 CCAAGTCTGAGTCCAGCAGTCTCCAGTGGGGCCAGCAGTCCAGCAGCAGCAGTCTGAGTCCA 1424
QY 264 -----ThrAsnGlyPheValSerSerTyThrIleHisIleLysAspAlaGlnValLysThr 282
DB 264 -----ThrAsnGlyPheValSerSerTyThrIleHisIleLysAspAlaGlnValLysThr 282

```

```

DB 1425 GCACCACTCTGAGTGGGGTCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGTCCAGT 1484
QY 282 lYAspThrValArgThrGlyGlnArgIleAlaSerMetLysAsnGlnProSerGlyValAla 302
DB 1485 GGGCTAGCAGCAGCAGCAGCAGTCTGAGTCCAGCAGCAGCAGCAGTCCAGTGGGGCCAGCAG 1544
QY 302 lAeuphGluPheArgIleSerArgAsnGlyVal 313
DB 1545 CCAACTCTGAGTCTGAGCAGCAGTCTGAGTCCAGTGGGATC 1579

```

RESULT 34
US-10-176-913-309
Sequence 309, Application US/10176913
Publication No. US2003002298A1
GENERAL INFORMATION:

APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jian
APPLICANT: Desnoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Guiney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Metanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
TITLE OF INVENTION: ACIDS ENCODING THE SAME
FILE REFERENCE: P34301C66
CURRENT APPLICATION NUMBER: US/10/176,913
CURRENT FILING DATE: 2002-06-20
Prior Application removed - See file wrapper or Palm
NUMBER OF SEQ ID NOS: 612
SEQ ID NO 309
LENGTH: 2436
TYPE: DNA
ORGANISM: Homo Sapien
US-10-176-913-309

Alignment Scores:
Pred. No.: 0.00289 Length: 2436
Score: 122.50 Matches: 78
Percent Similarity: 39.04% Conservative: 52
Best Local Similarity: 23.42% Mismatches: 145
Query Match: 7.34% Indels: 60
Gaps: 12

US-10-018-706-2 (1-322) x US-10-176-913-309 (1-2436)

```

QY 6 AAlaIleAsnSerGlnAsnGlnLysProIleYsArgLeuGlyLeuIlePheGlyValIle 25
DB 686 GCACCACTCTGAGTCCAGCAGCAGTCTGAGTCCAGTCCAGCAGCAGCAGCAGCAGCAG 745
QY 26 ThrThrCysIleLeuAlaGlyCysAlaSerLysProThrTyraSerThrSerGlySer 45
DB 746 TCTAGCAGCAGTCTCCAGTGGGGCC-----AGCAGCAGCAGCAGCAGCAGCAGCAGC 787
QY 46 GlySerHisArgThrSerGlySerGlyLeuAlaIleGlySerGlnValIleThrAsp 65
DB 788 GACTCCAGCAGCAGCAGTCCAGTGGGGCTGAGCAGCAGCAGCAGCAGTCTGAGTCCAGCAG 847
QY 66 SerGlnGlyValProAsnArgTyrgln-----ValIleGlnGlyAsp 79
DB 848 TCCAGTGGGGCCAG-CACAGCCAGCAGCAGTCTGAGTCCAGCAGCAGTCTGAGTGGGGCCAG 906
QY 80 ThrValSerLysIleAlaGlnArgTyrglyLeuAsnTrpArgGluIleGlyHis----- 97
DB 907 CACTGCAGCAGCAGTCTGAGTCCAGCAGCAGCAGTCTGAGTGGGGCCAGCAGCAGCAGCAG 966
QY 98 -----IleAsnAsnLeuAsnSerSerTyThrIleTyThrGlyGlnTrpLeuThrIle 115
DB 967 TGAAGTCCAGAGCAGTCTCCAGTGGGGCTGAGCAGCAGCAGCAGCAGTCTGAGTCCAGCAG 1026

```

QY	115	uTrpSerGIyAhrLeuLySValArg-----GIuArgSerTLeSerSerGIyValAs	132
Db	1027	STCCAGTGGGGCCAGCAGCCACCACCTGTGACTTCCAGACAGAGTGTCCAGTGGGGCCAG	1086
QY	132	nThrAla-HisThrProSerProValAlaValAlnSerSerArgProProVal-----G	150
Db	1087	CAGTCCCAACCAACTGTGAGTCCA-----GCACGACCTCCAGTGGGGCCAG	1133
QY	150	InGlnHisProAlaValAlnGlnLySProThrProProProValValValLySlySProThrP	170
Db	1132	CACAGCCCAACCACTGTGAGTCCAGCAGCACTCCAGTGGGGGTACACAGCCACCAACTTC	1191
QY	170	roThrProProValValAlnGlnInPro-----AlaProValAlaProProVal-----T	186
Db	1192	TGAGTCCCA-----GCACACCTCCAGTGGGGCCGCGACAGCCACCAACTGTGAGTCCAG	1244
QY	186	hrGIuAlaProPheAlaThrGlySerSerGIyValMetGlnPheArgTyPProValGIyA	206
Db	1246	CACAGTGTCCAGTGGGATCCAGCAGCAGTCAACCAATTCTGATGCCAGACACACCTCC-AGTG	1304
QY	206	laThrAsnProValValArgArgPheGlyThrAlaThValAlaGlySerThrValThrS	226
Db	1305	GGGGCCAAC-----ACAGCCACCACTGTGAGTCCAGTACGACTT	1343
QY	226	eTrAsnGIyMetTrPheSerSerGIyArgAspGIyAspLeuIleAsnAlaSerAsnAlaGIyT	246
Db	1344	CCAGTGGG-----GCCAACACAGCCA	1364
QY	246	hrValIleGlnAlaAspHisAsnMetAspArgIyAlaSerTLeValIleGlnHis-----	263
Db	1365	CCAACTGTGAGTCCAGCAGCAGTGTCCAGTGGGGCCAGACCTGCCACCAACTGTGAGTCCA	1422
QY	264	-----ThrAsnGIyPheValSerSerTyTLeHisIleLysAspAlaGlnValLyThrG	282
Db	1425	GCACACACTCCAGTGGGGGTCCAGACAGCCACCAACTGTGAGTCCAGCAGCAACTCCAGTGTG	1488
QY	282	lyAspThrValArgThrGlyGlnArgIleAlaSerMetLysAsnGlnProSerGIyAlaA	302
Db	1485	GGGCTAGACAGCCCAACTGTGATCCAGCAGCAACTCCAGTAGGCGAGCAGCAGCCA	1544
QY	302	laLeuPheGlnPheArgTLeSerArgAsnGIyVal	313
Db	1545	CCAACTGTGAGTCCAGCAGCAGTCCAGTGGGATC	1579

```

RESULT 35
US-10-180-552-309
; Sequence 309, Application US/10180552
; Publication No. US20030022300A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Utan
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P343081C153
; CURRENT APPLICATION NUMBER: US/10/180,552
; PRIORITY FILING DATE: 2002-06-25
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 309
; LENGTH: 2436
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-180-552-309

```

Alignment Scores:	
Pred. No.:	0.00289
Score:	122.50
Percent Similarity:	39.04%
Best Local Similarity:	23.42%
Query Match:	7.34%
DB:	9
Length:	2436
Matches:	78
Conservative:	52
Mismatches:	145
Indels:	60
Gaps:	12

```

US-10-018-706-2 (1-322) x US-10-180-552-309 (1-2436)
QY      6 AlaIleAsnSerGlnAsnGlnLysProIleLeuSarTyrLeuGlyLeuIlePheGlyValIle   25
Db      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
        686 GCCACCACTTCGTGAATCGAGCACAAGTGCATGGGGCCAGCACACTGCCAACCACTTGAG    745
QY      26 ThrThrCysIleLeuAlaGlyCysAlaSerLysProThrTyraSerThrSerylser     45
Db      |::::|::::|::::|::::|::::|::::|::::|::::|::::|::::|::::|::::|
        746 TCTMGACCACCTCTCCAGTAGGGGCC-----AGACGACCAACCAACTT    787
QY      46 GlySerHisArgThrSerylSerylGlyLeuAlaIleGlySerGlnValIleThrAsp     65
Db      |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
        788 GAGTTCCAGCACAACTCCAGTGGGGGCTGACAGACCAACCACTCGAATCCAGACCAAC    847
QY      66 SerGlnGlyValProAsnArgTyrGln-----ValylGlnGlyasp              79
Db      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
        848 TCcAGTGGGGCCAG-CACAGCCACCACTTCGATCCGACAGAAGTCCAGTAGGGCCAG    906
QY      80 ThrValSerLysIleAlaGlnArgTyrGlyLeuAsnTPArgGluIleGlyHis-----   97
Db      |::::|::::|::::|::::|::::|::::|::::|::::|::::|::::|::::|::::|
        907 CACTGCCACCACTTCGATCCAGACCACTTCAGTGGGGCCAGCACAGCCAACCACTC    966
QY      98 -----IleAsnAsnLeu-AasnSerTyr-ThrIleTyrThrGlyIntRpleuThrle     115
Db      |::::|::::|::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
        967 TGAGTCCAGAGCAAGCTTCATGTGGGGCTGGACAGCCACCAACTCGATCCAGACAGC    1022
QY      115 uTRPSerGlyAspLeuLysValArg-----GluARSerIleSerSerylValas       132
Db      |::::|::::|::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
        1027 CTCcAGTGGGGCCAGCACAGCCACCACTCGATCCGACAGTGTCCAGTGGGGCCAG    1086
QY      132 nThraAla-HisThr-ProSerProValAlaValGlnSerSerArgProProval-----g     150
Db      |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
        1087 CACTGCCACCACTTCGTGAATC-----GCACGACCTCGAGTGGGGCCAG    1133
QY      150 IngInhiSProAlaValGlnLysProThrProProvalValValLysProThrp         170
Db      |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
        1132 CACAGCCACCAACTCTGATGCACAGCAGCACTCCAGTGGGGCTGACAGCAGCAACCACTC    1199
QY      170 roTHrProProvalValGlnGlnPro--AlaproValAlaproProval-----t      186
Db      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
        1192 TGACTCCA-----GCACCAACTTCAGTGGGGCCGCGCACACCCCAACTCGATGCCAG    1245
QY      186 hrGluAlaProPhelaThrGlyserserylValMetGlnPheargyrrProvalGlya       206
Db      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
        1246 CACAGTGTCCAGTGGGATCAGACAGTACCATAATTCTAGTCCAGCACACCTTCC-AGTG    1304
QY      206 IatHrasnProvalValArGarphneGIThrAlathrValalaglyserThraVIThrs     226
Db      |::::|::::|::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
        1305 GGSGCCAA-----ACAGGCCACCACTCGAGTCCAGTACGACT    1343
QY      226 eraSnGlyMeTrPheserserylArGaSpGlyAspLeuIleAsnAlaserAsnAlaglyT     246
Db      |::::|::::|::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
        1344 CCAGTGGG-----GCCAACCAAGCA    1367
QY      246 hrvAlIleGlAlaAspHisAsnMetAspGlyAlaserIleValIleGlnhis-----      263
Db      |::::|::::|::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
        1365 cCAACTCGAATCCAGCAGAGTGTCCAGTGGGGCCAGCATGCCAACCACTCTGAGTCA    1422
QY      264 ----ThrAsnGlyPheValaserSeryTYrleniStileYAspAlaginVallystHrg     282
Db      |::::|::::|::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
        1425 GCACCAACTCCAGTGGGGTGCAGACAGCACCACCACTCGAATCCAGCAACAACCTCGAGT    1488
QY      282 lyASpThnValArgThrClglNarGllleAlaserMetLyVaSnGlnProSerylAlaa     302
Db      |::::|::::|::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
        1485 GGGCTAGAGCAGCCAACTTCGATCCCAAGTCCAGCAACCTCCAGTAGGGCCAGCAGCA    1544

```

QY 302 lAleupheglupheargileSerArgAsnGlyVal 313
 DB 1545 CCAACTCTGAGTCTAGACACAGTGTCCAGTGGGATC 1579
 RESULT 36
 US-10-180-557-309
 / Sequence 309, Application US/10180557
 / Publication No. US20030022301A1
 / GENERAL INFORMATION:
 / APPLICANT: Baker, Kevin P.
 / APPLICANT: Chen, Jian
 / APPLICANT: Desnoyers, Luc
 / APPLICANT: Goddard, Audrey
 / APPLICANT: Godowski, Paul J.
 / APPLICANT: Gurney, Austin L.
 / APPLICANT: Pan, James
 / APPLICANT: Smith, Victoria
 / APPLICANT: Watanabe, Colin K.
 / APPLICANT: Wood, William I.
 / APPLICANT: Zhang, Zemin
 / TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
 / ACIDS ENCODING THE SAME
 / FILE REFERENCE: P3430R1C147
 / CURRENT APPLICATION NUMBER: US/10/180,557
 / CURRENT FILING DATE: 2002-06-25
 / Prior Application removed - See File Wrapper or Palm
 / NUMBER OF SEQ ID NOS: 612
 / SEQ ID NO 309
 / LENGTH: 2436
 / TYPE: DNA
 / ORGANISM: Homo Sapien
 / US-10-180-557-309
 Alignment Scores:
 Pred. No.: 0.00289 Length: 2436
 Score: 122.50 Matches: 78
 Percent Similarity: 39.04% Conservative: 52
 Best Local Similarity: 23.42% Mismatches: 145
 Query Match: 7.34% Indels: 60
 DB: 9 Gaps: 12
 US-10-018-706-2 (1-322) x US-10-180-557-309 (1-2436)
 QY 6 AlaIleAsnSerGlnAsnGlnIysProIleYsArgLeuGlyLeuIlePheGlyValIle 25
 DB 686 GCACCACTCTGAGTCTAGACACAGTGTCCAGTGGGCGACAGCTGCCCACTCTGAG 745
 QY 26 ThrThrcyaleuIleuIaGlyCysAlaSerIysProThrIysAsnSerThrseryler 45
 DB 746 TCTAGCACACTCTCCAGTGGGCGCC-----AGCACGCCACCACTCT 787
 QY 46 GlySerHisArgThrserylerGlyGlyLeuAlaIleGlySerGlnValIleThrsP 65
 DB 788 GACTCCAGCACCAACTCTGAGTGGGCGTACACAGCCCACTCTGAGTCCAGCACCAACC 847
 QY 66 SerGlnGlyValProAsnArgIysGln-----ValIleGlnIysP 79
 DB 848 TCCAGTGGGCGCCAG-CACAGCCACCAACTCTGAGTCCAGCACAGTGTCCAGTGGGCGCAG 906
 QY 80 ThrValSerIysIleAlaGlnArgIysGlyLeuAsnTrpArgGlnIleGlyHis----- 97
 DB 907 CACTGCCACCACTCTGAGTCCAGCACCACTCTGAGTGGGCGCGACAGCCACCACTCTC 966
 QY 98 -----IleAsnAsnIeu-AsnSerSerIysTrhIleIysTrhGlyGlnTrpIleThrIle 115
 DB 967 TGAAGTCCAGAACGACCTCCAGATGGGCGTGGCGACAGCCACCACTCTGAGTCCAGCACAGC 1026
 QY 115 uTPSerGlyAspLeuIysValArg-----GluArgSerIleSerSerGlyValAs 132
 DB 1027 CTCGATGTGGGCGACAGCACAGCCCAACTCTGAGTCCAGCACAGTGTCCAGTGGGCGCAG 1086
 QY 132 nThrAla-HisThrProSerProValAlaValGlnSerSerArgProProVal-----G 150

DB 1087 CACTGCCACCACTCTGAGTCTCA-----GCAGGACCTCCAGTGGGCGCAG 1131
 QY 150 InGlnIisProAlaValGlnIysProThrProProValValValIysIysProThrP 170
 DB 1132 CACAGCCACCACTCTGAGTCCAGCACAGCACTCTCAGTGGGCGTACAGCACCACTCTC 1191
 QY 170 roThrProProValValGlnGlnIysPro-----AlaProValAlaProProVal-----T 186
 DB 1192 TGAATCCA-----GCACCACTCTCCAGTGGGCGCGCGACAGCCACCACTCTGAGTCCAG 1245
 QY 186 hrGlnAlaProPheIleThrIysSerSerGlyValMetGlnPheArgIysTrpProValGly 206
 DB 1246 CACAGTGTCCAGTGGGATGACAGCACTCCCAATTTCTGATGTCACACACCTCC-AGTG 1304
 QY 206 lArhAsnProValValArgAspPheGlyThrAlaThrValAlaGlySerThrValThrs 226
 DB 1305 GGGCCAAAC-----ACAGCCACCACTCTGAGTCCAGTCCAGTCT 1343
 QY 226 exArgnGlyMetTrpPheSerGlyArgAspGlyAspLeuIleAsnAlaSerAsnIleGlyT 246
 DB 1344 CAGTGGG-----GCACACAGCCA 1364
 QY 246 hrValIleGlnAlaAspHisAsnMetAspGlyAlaSerIleValIleGlnHis----- 263
 DB 1365 CCAACTCTGAGTCCAGCACAGTGTCCAGTGGGCGCGACAGCTGCCACCACTCTGAGTCCA 1424
 QY 264 -----ThrAsnGlyPheValSerSerIysTrhIleHisIleIysAspAlaGlnValIysThrg 282
 DB 1425 GCACCAACTCTGAGTGGGCGTACAGCACAGCCACCACTCTGAGTGGCGACCACTCTCAGTG 1484
 QY 282 lYAspThrValArgThrGlyGlnArgIleAlaSerMetIysAsnGlnProSerGlyAla 302
 DB 1485 GGGCTGACACAGCCACCACTCTGAGTCCAGCACCACTCTCCAGTGGGCGACAGCCA 1544
 QY 302 lAleupheglupheargileSerArgAsnGlyVal 313
 DB 1545 CCAACTCTGAGTCTAGACACAGTGTCCAGTGGGATC 1579
 RESULT 37
 US-10-173-700-309
 / Sequence 309, Application US/10173700
 / Publication No. US20030027262A1
 / GENERAL INFORMATION:
 / APPLICANT: Baker, Kevin P.
 / APPLICANT: Chen, Jian
 / APPLICANT: Desnoyers, Luc
 / APPLICANT: Goddard, Audrey
 / APPLICANT: Godowski, Paul J.
 / APPLICANT: Gurney, Austin L.
 / APPLICANT: Pan, James
 / APPLICANT: Smith, Victoria
 / APPLICANT: Watanabe, Colin K.
 / APPLICANT: Wood, William I.
 / APPLICANT: Zhang, Zemin
 / TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
 / ACIDS ENCODING THE SAME
 / FILE REFERENCE: P3430R1C14
 / CURRENT APPLICATION NUMBER: US/10/173,700
 / CURRENT FILING DATE: 2002-06-17
 / Prior Application removed - See File Wrapper or Palm
 / NUMBER OF SEQ ID NOS: 612
 / SEQ ID NO 309
 / LENGTH: 2436
 / TYPE: DNA
 / ORGANISM: Homo Sapien
 / US-10-173-700-309
 Alignment Scores:
 Pred. No.: 0.00289 Length: 2436
 Score: 122.50 Matches: 78
 Percent Similarity: 39.04% Conservative: 52
 Best Local Similarity: 23.42% Mismatches: 145

Query Match: 7.34% Indels: 60
 DB: 9 Gaps: 12
 US-10-018-706-2 (1-322) x US-10-173-700-309 (1-2436)

QY 6 AlaIleAnSerGlnAnGlnLysProIleYAlaArgLeuGlyLeuIlePheGlyValIle 25
 DB 686 GCCACCAACTCTGAGTCAGACAGAGTCAGTAGGGCCAGCACTGCCACCAACTCTGAG 745
 QY 26 ThrThrCysIleLeuAlaGlyCysAlaSerIleProThrIleYAlaSerIleThrSer 45
 DB 746 TCTAGCCACCTCTCAGTGGGGCC-----ACAGCCAGCCCAACTCT 787
 QY 46 GlySerHisArgThrSerGlySerGlyGlyLeuAlaIleGlySerGlnValIleThrAsp 65
 DB 788 GACTCCAGCAACAACCTCAGTGGGGCTAGACAGCCCAACTCTGAGTCAGACCAACC 847
 QY 66 SerGlnGlyValProAsnArgTyrgln-----VallySerGlnGlyAsp 79
 DB 848 TCCAGTGGGGCCAG-CACAGCCCAACTCTGAGTCAGACAGAGTCAGTAGGGCCAG 906
 QY 80 ThrValSerIleAlaGlnArgTyrglyLeuAsnTrpArgGluIleGlyHis----- 97
 DB 907 CACTGCCACCAACTCTGAGTCAGACCAACTCTCAGTGGGGCCAGCAAGCCCAACTC 966
 QY 98 -----IleAsnAsnLeu-AsnSerSerTyrglyThrIleTyrglyThrGlyGlnTrpLeuThr 115
 DB 967 TGAATCCAGAACGACCTCTCAATGGGGCTGGACAGCCCAACTCTGAGTCAGACCAAC 1026
 QY 115 uTrpSerGlyAspLeuLysValArg-----GluArgSerIleSerSerGlyValAs 132
 DB 1027 CTCAGTGGGGCCAGCAAGCCCAACTCTGAGTCAGACAGAGTCAGTAGGGCCAG 1086
 QY 132 nThrAla-HisThrProSerProValAlaValGlnSerSerArgProProVal-----G 150
 DB 1087 CACTGCCACCAACTCTGAGTC-----GCAGCACTCTCAGTGGGGCCAG 1131
 QY 150 lngInHisProAlaValGlnLysProThrProProValValValLysProThr 170
 DB 1132 CACAGCCCAACTCTGAGTCAGACCAACTCTCAGTGGGGCTAGACAGCCCAACTCTC 1191
 QY 170 roThrProProValValGlnGlnPro--AlaProValAlaProProVal-----T 186
 DB 1192 TGAATCCA-----GCACCACTCTCAGTGGGGCCGGCCAGCAAGCACTCTGAGTCAG 1245
 QY 186 hrgIuAlaProPheAlaThrGlySerSerGlyValMetGlnPheArgTyrglyProValGly 206
 DB 1246 CACAGTCTCAGTGGGGATCAGACAGATCCCAATCTGAGTCAGACCACTCTC-AGTG 1304
 QY 206 laThrAsnProValValArgArgPheGlyThrAlaThrValAlaGlySerThrValThrs 226
 DB 1305 GGGCCAC-----ACAGCCCAACTCTGAGTCAGTCAGTCAGACT 1343
 QY 226 eraAnglyMetTrpPheSerGlyArgAspGlyAspLeuIleAsnAlaSerAsnAlaGly 246
 DB 1344 CCAGTGGG-----GCCAACACAGCA 1364
 QY 246 hrValIleGlnAlaAspHisAsnMetAspGlyAlaSerIleValIleGlnHis----- 263
 DB 1365 CCAACTCTGAGTCAGACAGATCTCAGTGGGGCCAGACAGCCCAACTCTGAGTC 1424
 QY 264 -----ThrAsnGlyPheValSerSerTyrglyHisIleLeuAspAlaGlnValLysThr 282
 DB 1425 GCACCACTCTCAGTGGGGCTAGACAGCCCAACTCTGAGTCAGACCAACTCTCAGTGG 1484
 QY 282 lyaSerThrValArgThrGlyGlnArgIleAlaSerMetLysAsnGlnProSerGlyValAs 302
 DB 1485 GGGGTAGACAGCAAGCAACTCTGAGTCAGACCAACTCTCAGTGGGGCCAGCAAGCA 1544
 QY 302 laLeuPheGlnPheArgIleSerArgAsnGlyVal 313
 DB 1545 CCAACTCTGAGTCAGACAGATCTCAGTGGGATC 1579

RESULT 38
 US-10-174-572-309
 : Sequence 309, Application US/10174572
 : Publication No. US20030027263A1
 : GENERAL INFORMATION:
 : APPLICANT: Baker, Kevin P.
 : APPLICANT: Chen, Jian
 : APPLICANT: Desnoyers, Luc
 : APPLICANT: Goddard, Audrey
 : APPLICANT: Godowski, Paul J.
 : APPLICANT: Gunney, Austin L.
 : APPLICANT: Pan, James
 : APPLICANT: Smith, Victoria
 : APPLICANT: Watanabe, Colin K.
 : APPLICANT: Wood, William I.
 : APPLICANT: Zhang, Zemin
 : TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
 : FILE REFERENCE: P3430R1C40
 : CURRENT APPLICATION NUMBER: US/10/174,572
 : PRIORITY FILING DATE: 2002-06-18
 : Prior Application removed - See File Wrapper or Paim
 : NUMBER OF SEQ ID NOS: 612
 : SEQ ID NO 309
 : LENGTH: 2436
 : TYPE: DNA
 : ORGANISM: Homo Sapien
 : US-10-174-572-309

Alignment Scores:
 Pred. No.: 0.00289 Length: 2436
 Score: 122.50 Matches: 78
 Percent Similarity: 39.04% Conservative: 52
 Best Local Similarity: 23.42% Mismatches: 145
 Query Match: 7.34% Indels: 60
 DB: 9 Gaps: 12

US-10-018-706-2 (1-322) x US-10-174-572-309 (1-2436)

QY 6 AlaIleAnSerGlnAnGlnLysProIleYAlaArgLeuGlyLeuIlePheGlyValIle 25
 DB 686 GCCACCAACTCTGAGTCAGACAGAGTCAGTAGGGCCAGCACTGCCACCAACTCTGAG 745
 QY 26 ThrThrCysIleLeuAlaGlyCysAlaSerIleProThrIleYAlaSerIleThrSer 45
 DB 746 TCTAGCCACCTCTCAGTGGGGCC-----ACAGCCAGCCCAACTCT 787
 QY 46 GlySerHisArgThrSerGlySerGlyGlyLeuAlaIleGlySerGlnValIleThrAsp 65
 DB 788 GACTCCAGCAACAACCTCAGTGGGGCTAGACAGCCCAACTCTGAGTCAGACCAACC 847
 QY 66 SerGlnGlyValProAsnArgTyrgln-----VallySerGlnGlyAsp 79
 DB 848 TCCAGTGGGGCCAG-CACAGCCCAACTCTGAGTCAGACAGAGTCAGTAGGGCCAG 906
 QY 80 ThrValSerIleAlaGlnArgTyrglyLeuAsnTrpArgGluIleGlyHis----- 97
 DB 907 CACTGCCACCAACTCTGAGTCAGACCAACTCTCAGTGGGGCCAGCAAGCCCAACTC 966
 QY 98 -----IleAsnAsnLeu-AsnSerSerTyrglyThrIleTyrglyThrGlyGlnTrpLeuThr 115
 DB 967 TGAATCCAGAACGACCTCTCAATGGGGCTGGACAGCCCAACTCTGAGTCAGACCAAC 1026
 QY 115 uTrpSerGlyAspLeuLysValArg-----GluArgSerIleSerSerGlyValAs 132
 DB 1027 CTCAGTGGGGCCAGCAAGCCCAACTCTGAGTCAGACAGAGTCAGTAGGGCCAG 1086
 QY 132 nThrAla-HisThrProSerProValAlaValGlnSerSerArgProProVal-----G 150
 DB 1087 CACTGCCACCAACTCTGAGTC-----GCAGCACTCTCAGTGGGGCCAG 1131
 QY 150 lngInHisProAlaValGlnLysProThrProProValValValLysProThr 170

Db 1132 CACAGCCACCACTCTGATCCAGCAGCACTCCAGTGGGGCTAGCAGCAGCCACCACTC 1191
 QY 170 rothProProvalValGlnGlnPro--AlaProValAlaProProVal-----T 186
 Db 1192 TGACTCCA-----GCACAACCTCCAGTGGGGCGCCAGCAGCCACCACTCTGATCCAG 1245
 QY 186 hrgLualProPhaIaThrgIySerSergIyValMetGlnPhaIaThrgIyProValIgiYa 206
 Db 1246 CACAGTCTCAGGGAGTCCAGCAGCTCCCAATCTGATGTCAGCAGCAGCCCTCC-AGTG 1304
 QY 206 lathrasProValValArgArgPheGlyThrAlaThraValAlaGlySerThrValThrs 226
 Db 1305 GGGCCCAAC-----ACAGCCCAACCTCTGATCTCAGTCCAGTCCAGTCC 1343
 QY 226 eThrasnglyMetTrpPheSergIyArgAspGlyAspLeuIleAsnIleSeraAsnAlaGlyT 246
 Db 1344 CCAGTGGG-----GCCAACACAGCCCA 1364
 QY 246 hrvAlIleGlnAlaAspHisAsnMetAspGlyAlaSerIleValIleGlnHis----- 263
 Db 1365 CCAACTCTGATCCACAGAGTGTCCAGTGGGGCCAGCTCCAGCAGCAACTCTGATCCCA 1424
 QY 264 -----ThrasnglyPheValSerSergIyIleHisIleIleAspAlaGlnValIySthrg 282
 Db 1425 GCACAACCTCCAGTGGGGTCCAGCAGCCACCACTCTGATCCAGCAGCAACCTCCAGTGG 1484
 QY 282 lYasPthraValArgThrgIyGlnArgIleAlaSerMetIleAsnGlnProSergIyAla 302
 Db 1485 GGGCTAGCAGAGCCCAACTCTGATCCAGCAGCAACTCTCCAGTGGGCGCAGCAGCA 1544
 QY 302 lAleupheGluPheArgIleSeraArgAsnGlyVal 313
 Db 1545 CCAACTCTGATCTAGCAGCAGTGTCCAGTGGGATC 1579
 RESULT 39
 US-10-174-579-309
 ; Sequence 309, Application US/10174579
 ; Publication No. US20030027264A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Baker, Kevin P.
 ; APPLICANT: Chen, Jian
 ; APPLICANT: Desnoyers, Luc
 ; APPLICANT: Goddard, Audrey
 ; APPLICANT: Godowski, Paul J.
 ; APPLICANT: Gurney, Austin L.
 ; APPLICANT: Smith, Victoria
 ; APPLICANT: Watanabe, Colin K.
 ; APPLICANT: Wood, William I.
 ; APPLICANT: Zhang, Zemin
 ; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
 ; FILE REFERENCE: P3430R/C3
 ; CURRENT APPLICATION NUMBER: US/10/174,579
 ; PRIOR APPLICATION REMOVED - See File Wrapper or Palm
 ; NUMBER OF SEQ ID NOS: 612
 ; SEQ ID NO 309
 ; LENGTH: 2436
 ; TYPE: DNA
 ; ORGANISM: Homo Sapien
 US-10-174-579-309
 Alignment Scores:
 Pred. No.: 0.00289 Length: 2436
 Score: 122.50 Matches: 78
 Percent Similarity: 39.04% Conservative: 52
 Best Local Similarity: 23.42% Mismatches: 145
 Query Match: 7.34% Indels: 60
 Db: 9 Gaps: 12

US-10-018-706-2 (1-322) x US-10-174-579-309 (1-2436)
 ; Sequence 309, Application US/10174582
 ; Publication No. US20030027265A1
 ; GENERAL INFORMATION:

QY 6 AlaIleAsnSergIleAsnGlnIlyProIleIySargIleGlyIleIlePheGlyValIle 25
 Db 686 GCCACCACTCTGATCCAGCAGCAGTGTCCAGTGGGGCGCCAGCAGCCACCACTCTGAG 745
 QY 26 ThrThrcysIleLeuAlaGlyCysAlaSerIlyProThryIyAsnSerThrsGlySer 45
 Db 746 TCTAGCACACTCTCCAGTGGGGCC-----AGCAGCCCAACCAACTCT 787
 QY 46 GlySerHisArgThrsSergIySergIyGlyLeuAlaIleGlySergIyValIleThrasp 65
 Db 788 GACTCCAGCAGCAACTCTCCAGTGGGGCTAGCAGCAGCCACCAACTCTGATCCAGCAGCAACC 847
 QY 66 SerGlnGlyValProAsnArgTyrgln-----ValIyGlnGlyAsp 79
 Db 848 TCCAGTGGGGCCAG-CACAGCCCAACTCTGATCCAGCAGCAGTGTCCAGTGGGGCCAG 906
 QY 80 ThrValSerIleIleAlaGlnArgTyrglyLeuAsnTPArgGlnIleGlyHis----- 97
 Db 907 CACTGCCACCACTCTGATCCAGCAGCAACTCTCCAGTGGGGCGCCAGCAGCAGCAACTC 966
 QY 98 -----IleAsnAsnLeu-AsnSerSergIyThrIleTyThrGlyIleIlePheIle 115
 Db 967 TGAGTCCAGAGCAGACTCCCAATGGGGCTGGCAGCAGCCCAACTCTGATCCAGCAGCAGC 1026
 QY 115 utrPserGlyAspLeuIySvalArg-----GlnArgSerIleSeraSergIyValAs 132
 Db 1027 CTCAGTGGGGCGCCAGCAGCAGCAACTCTGATCCAGCAGCAGTGTCCAGTGGGGCGCCAG 1086
 QY 132 nThraIle-HisThrProSeraProValAlaValAlaGlnSeraArgProProVal-----G 150
 Db 1087 CACTGCCACCAACTCTGATCCCA-----GCCAGCACTCTCAGTGGGGCGCCAG 1131
 QY 150 IlnGlnHisProAlaValAlaGlnIlySProThrProProValValValIyIlySProThrP 170
 Db 1132 CACAGCCCACTCTGATCCAGCAGCAGCAGTGTCCAGTGGGGCTAGCAGCAGCCCAACTC 1191
 QY 170 rothProProValValGlnGlnPro--AlaProValAlaProProVal-----T 186
 Db 1192 TGACTCCA-----GCACAACCTCCAGTGGGGCGCCAGCAGCCACCACTCTGATCCAG 1245
 QY 186 hrgLualProPhaIaThrgIySerSergIyValMetGlnPhaIaThrgIyProValIgiYa 206
 Db 1246 CACAGTCTCAGGGAGTCCAGCAGCTCCCAATCTGATGTCAGCAGCAGCCCTCC-AGTG 1304
 QY 206 lathrasProValValArgArgPheGlyThrAlaThraValAlaGlySerThrValThrs 226
 Db 1305 GGGCCCAAC-----ACAGCCCAACCTCTGATCTCAGTCCAGTCCAGTCC 1343
 QY 226 eThrasnglyMetTrpPheSergIyArgAspGlyAspLeuIleAsnIleSeraAsnAlaGlyT 246
 Db 1344 CCAGTGGG-----GCCAACACAGCCCA 1364
 QY 246 hrvAlIleGlnAlaAspHisAsnMetAspGlyAlaSerIleValIleGlnHis----- 263
 Db 1365 CCAACTCTGATCCACAGAGTGTCCAGTGGGGCGCCAGCTCCAGCAGCAACTCTGATCCCA 1424
 QY 264 -----ThrasnglyPheValSerSergIyIleHisIleIleAspAlaGlnValIySthrg 282
 Db 1425 GCACAACCTCCAGTGGGGTCCAGCAGCCACCACTCTGATCCAGCAGCAACCTCCAGTGG 1484
 QY 282 lYasPthraValArgThrgIyGlnArgIleAlaSerMetIleAsnGlnProSergIyAla 302
 Db 1485 GGGCTAGCAGAGCCCAACTCTGATCCAGCAGCAACTCTCCAGTGGGCGCAGCAGCA 1544
 QY 302 lAleupheGluPheArgIleSeraArgAsnGlyVal 313
 Db 1545 CCAACTCTGATCTAGCAGCAGTGTCCAGTGGGATC 1579
 RESULT 40
 US-10-174-582-309
 ; Sequence 309, Application US/10174582
 ; Publication No. US20030027265A1
 ; GENERAL INFORMATION:

APPLICANT: Baker, Kevin P.
 APPLICANT: Chen, Jian
 APPLICANT: Desnoyers, Luc
 APPLICANT: Goddard, Audrey
 APPLICANT: Godowski, Paul J.
 APPLICANT: Guiney, Austin L.
 APPLICANT: Pan, James
 APPLICANT: Smith, Victoria
 APPLICANT: Watanabe, Colin K.
 APPLICANT: Wood, William I.
 APPLICANT: Zhang, Zemin
 TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
 FILE REFERENCE: P3430R1C36
 CURRENT APPLICATION NUMBER: US/10/174,582
 CURRENT FILING DATE: 2002-06-18
 Prior Application removed - See File Wrapper or Palm
 NUMBER OF SEQ ID NOS: 612
 SEQ ID NO 309
 LENGTH: 2436
 TYPE: DNA
 ORGANISM: Homo Sapien
 US-10-174-582-309

Alignment Scores:
 Pred. No.: 0.00289 Length: 2436
 Score: 122.50 Matches: 78
 Percent Similarity: 39.04% Conservative: 52
 Best Local Similarity: 23.42% Mismatches: 145
 Query Match: 7.34% Indels: 60
 DB: 9 Gaps: 12
 US-10-018-706-2 (1-322) x US-10-174-582-309 (1-2436)

QY 6 AlaIleAsnSerGlnAsnGlnProIleValGluGlyLeuIlePheGlyValIle 25
 DB 686 GCCACCACTCTGAGTCCAGACAGTGTCCAGTGGCCAGACACCTCCAGCACTCTGAG 745
 QY 26 ThrThrGlySileuAlaGlyCyAlaSerLysProThrTyraAsnSerThrsGlySer 45
 DB 746 TCTAGCACACTCTCCAGTGGGCC-----AGCACGCCACCACTCT 787
 QY 46 GlySerHisArgThrsSerGlySerglyLeuAlaIleGlySerglnValIleThrsAsp 65
 DB 788 GACTCCAGCACCACTCTCCAGTGGGGCTAGCACAGCCCACTCTGAGTCCAGCACAC 847
 QY 66 SerGlnGlyValProAsnArgTyrlin-----ValLysGlnGlyAsp 79
 DB 848 TCCAGTGGGGCCAG-CACAGCCACCACTCTGAGTCCAGACAGTGTCCAGTGGGCCAG 906
 QY 80 ThrValSerLysIleAlaGlnArgTyrlGlyLeuAsnTrpArgGlnIleGlyHis----- 97
 DB 907 CACTGCCACCACTCTGAGTCCAGACCACTCTCCAGTGGGCCAGCACAGCCCACTCTC 966
 QY 98 -----IleAsnAsnLeu-AsnSerSerTyrlleTyrlleThrsGlnTrpLeuThrsIle 115
 DB 967 TGAAGTCAGAAAGCACTCTCCAGTGGGGCTGGCACACCACTCTGAGTCCAGCACAG 1026
 QY 115 uTrpSerGlyAspLeuLysValArg-----GluArgSerIleSerSerGlyValAs 132
 DB 1027 CTCCAGTGGGGCCAGCACCACTCTGAGTCCAGACAGTGTCCAGTGGGGCCAG 1086
 QY 132 nThraIle-HisThrProSerProValAlaValGlnSerSerArgProVal-----G 150
 DB 1087 CACTGCCACCACTCTGAGTCCA-----GCACGACCTCCAGTGGGGCCAG 1131
 QY 150 lnglnHisProAlaValAlaGlnLysProThrProProValAlaValAlaLysLysProThrP 170
 DB 1132 CACAGCCACCACTCTGAGTCCAGACAGCACTCTCCAGTGGGGCTAGCACAGCCCACTCTC 1191
 QY 170 roThrProProValValGlnGlnPro---AlaProValAlaProProVal-----T 186
 DB 1192 TGACTTCCA-----GCACCAACTCTCCAGTGGGGCCGGCACAGCCCACTCTGAGTCCAG 1245

QY 186 hrGluAlaProPheAlaThrGlySerSerGlyValMetGlnPheArgTyProValGlyA 206
 DB 1246 CACAGTGTCCAGTGGGGATTCAGACAGTCCACCAATTCTGAGTCCAGCACACCTCTC-AGTG 1304
 QY 206 lArThrAsnProValValArgArgPheGlyThraIleThraValAlaGlySerThraIleThrs 226
 DB 1305 GGGCCCAAC-----ACAGCCACCACTCTGAGTCCAGTCCAGTCCAGTCC 1343
 QY 226 eArgnGlyMetTrpPheSerGlyValArgAspGlyAspLeuIleAsnAlaSerAsnAlaGlyT 246
 DB 1344 CCAAGTGG-----GCCAACACAGCCA 1364
 QY 246 hrValIleGlnAlaAspHisAsnMetAspGlyAlaSerIleValIleGlnHis----- 263
 DB 1365 CCAACTGTGAGTCCAGACAGTGTCCAGTGGGGCCAGACACTGCCACCACTCTGAGTCCA 1424
 QY 264 -----ThraGnGlyPheValSerSerTyrlleHisIleLysAspAlaGlnValLysThrg 282
 DB 1425 GCACCACTCTCCAGTGGGGTCCAGCACAGCCCACTCTGAGTCCAGCACCAACTCTCAGTG 1484
 QY 282 lYAspThrValArgThrsGlyGlnArgIleAlaSerMetLysAsnGlnProSerGlyAla 302
 DB 1485 GGGTACACAGCCCACTCTGAGTCCAGACCACTCTCCAGTGGGCCAGCACAGCCA 1544
 QY 302 lAlaLeuPheGlnPheArgIleSerArgAsnGlyVal 313
 DB 1545 CCAACTCTGAGTCCAGACAGTGTCCAGTGGGATC 1579

Search completed: July 4, 2003, 00:39:18
 Job time : 8989 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 3, 2003, 18:36:07 ; Search time 1135 Seconds

(without alignments)
4594.667 Million cell updates/sec

Title: US-10-018-706-2

Perfect score: 1670

Sequence: 1 MTVTAINSONOKPIKRLGL.....LFEFRISRNQYVDPITVLK 322

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5

Xgapop 10.0 , Xgapext 0.5

Xgapop 6.0 , Xgapext 7.0

Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n model -DEV=x1h

-Q=/cgn2_1/USFPO.spool/US10018706/runat_30062003_091105_23805/app_query.fasta_1.519

-DB=EST -QFMT=fastap -SUFFIX=ret -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0

-UNITS=bits -START=1 -END=1 -MATRIX=Dlosum62 -TRANS=human40.cdi -LIST=45

-DOCLALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=40 -MODE=LOCAL

-OUTFMT=ptc -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000

-USER=US10018706 @CGN 1.1 1456 @runat_30062003_091105_23805 -NCPU=6 -ICPU=3

-NO_MMAR -LARGOQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG

-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6

-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST :

1: em_estba.*

2: em_esthum.*

3: em_estin.*

4: em_estmu.*

5: em_estov.*

6: em_estpl.*

7: em_estro.*

8: em_hlc.*

9: gb_est1.*

10: gb_est2.*

11: gb_hlc.*

12: gb_est3.*

13: gb_est4.*

14: gb_est5.*

15: em_esthum.*

16: em_estom.*

17: gb_gss.*

18: em_gss_hum.*

19: em_gss_inv.*

20: em_gss_pln.*

21: em_gss_vrt.*

22: em_gss_fun.*

23: em_gss_mam.*

24: em_gss_mus.*

25: em_gss_other.*

26: em_gss_pro.*

27: em_gss_rtd.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	139.5	8.4	784	12	BF866446
2	136	8.1	906	17	CNS01DTR
3	134	8.0	4812	14	BH770991
4	132	7.9	433	14	BQ791034
5	132	7.9	1096	17	AF094939
6	131	7.8	596	12	BG543924
7	131	7.8	1306	14	BM921488
8	129.5	7.8	398	14	BQ791640
9	127	7.6	890	17	CNS03DTR
10	126.5	7.6	669	14	BQ704548
11	125.5	7.5	628	10	AV520024
12	125.5	7.5	1279	14	BM911051
13	124.5	7.5	666	12	BG593928
14	124.5	7.5	767	14	BQ505542
15	124.5	7.5	839	17	BF859163
16	124	7.4	285	17	BH705150
17	124	7.4	735	17	AZ187441
18	124	7.4	1252	14	BM913728
19	123.5	7.4	538	10	BE240905
20	123.5	7.4	1002	17	CNS018ZG
21	123	7.4	689	13	BQ220947
22	122.5	7.3	500	9	AU268480
23	122.5	7.3	722	17	AG130154
24	122.5	7.3	1363	17	AG054509
25	122	7.3	450	13	BM109778
26	121.5	7.3	1140	17	AG159588
27	120.5	7.2	508	13	BI174098
28	120.5	7.2	969	17	CNS03NCR
29	120	7.2	365	9	AL377972
30	120	7.2	511	10	AW568923
31	120	7.2	650	14	BQ147993
32	120	7.2	710	13	BI217940
33	120	7.2	1460	13	BM457264
34	119	7.1	567	12	BG607114
35	119	7.1	909	14	BQ646351
36	118	7.1	443	9	AI484465
37	118	7.1	443	10	BE460175
38	117.5	7.0	615	14	H76194
39	117.5	7.0	329	14	BH244723
40	117	7.0	754	14	U74096
41	117	7.0	586	10	AM033946
42	117	7.0	619	10	AV520316
43	117	7.0	864	17	AZ183812
44	116.5	7.0	717	17	AG132838
45	116.5	7.0	772	13	BM050011

ALIGNMENTS

RESULT 1
BF866446/c
LOCUS
DEFINITION 963069A11.x1 C. reinhardtii CC-1690, Stress condition I, normalized
lambda zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION BF866446
VERSION BF866446.1 GI:12256590
KEYWORDS EST.
SOURCE Chlamydomonas reinhardtii.
ORGANISM Chlamydomonas reinhardtii.
REFERENCE 1 (bases 1 to 784)
Lefebvre, P., McDermott, J. P., Shragar, J., Silflow, C. and Stern, D.

TITLE
Analyses of the Chlamydomonas reinhardtii Genome: A Model,
Unicellular System for Analyzing Gene Function and Regulation in
Vascular Plants; project phase 3

JOURNAL
Contact: Charles Hauser
DCMB Box 91000
Durham, NC 27706-1000
Tel.: 919 613 8159
Fax: 919 613 8177
Email: chauer@cduhe.edu.

FEATURES
Location/Qualifiers
1..784
/organism="Chlamydomonas reinhardtii"
/strain="CC-1690 wild type mc+ 21gr"
/db_xref="taxon:3055"
/clone_1lb="C. reinhardtii CC-1690, stress condition I,
normalized, lambda zap ii"
/note="vector: plusescript II SK-, site 1: EcoRI; Site 2:
XhoI; This library, constructed by John Davies and Jeffrey
McDermott, combines cDNAs from CC-1690 cells grown to
mid-log phase in TAP-N (30 min, 1hr, 4hr), TAP-S (30 min,
1hr, 4hr), TAP-P (4hr, 12hr, 24hr), NO3 to NH4 (30min, 1hr
, 4hr) and NH4 to NO3 (30min, 1hr, 4hr). PolyA mRNA was
purified from each sample, pooled and cDNA synthesized.
The cDNA was directionally cloned into lambda zap II.
(Stratagene) in the EcoRI (5') and XhoI (3') sites.
Plusescript II SK- plasmids were excised from the lambda
zap clones by superinfection with Exbsist (Stratagene)
phase. The library was normalized using method 4 described
in Bonaldo et al (1996) Genome Research 6: 791-806."

BASE COUNT
171 a 244 c 196 g 169 t 4 others

ORIGIN
Alignment Scores:
Pred. No.: 0.0128 Length: 784
Score: 139.50 Matches: 61
Percent Similarity: 40.27% Conservative: 30
Best Local Similarity: 26.59% Mismatches: 71
Query Match: 8.35% Indels: 65
DB: 12 Gaps: 10

US-10-018-706-2 (1-322) x BF866446 (1-784)

Qy	62	ValllStThrAspSerGlnGlyValProAsnAsgTyrGlnValIleGlnGlnGlyAspThrVal	81
Db	595	ATATACACCCTGCAGCGACTTGCACTTCCTTCGACCTGGCGGTGGCGAAAGATTACTATA	536
Qy	82	SerLysAlaIleGlnAgtTyrGlyLeuAsnTPPrAgcLuIllegLyHis-----	97
Db	535	AGTTCTTACTCT---CGCTCAATATATGGGTGGGGGGATTAAGCAAGAAGACCCGAAG	479
Qy	98	-----IleAsnLeuAsnSer	104
Db	478	ACCTGCACCTCATCTTAGTCAGCAGCGTATATTAGTGTAGACAGAAACAATATCAGAGGGG	419
Qy	105	TyrTrlletYrrTrngIynIntPleuthLeu-----TrpserGlyAspLeuVal	122
Db	418	AACACATGACAAAGAAAGAAATAGTTACCGTCCGTTGGTACTCTCTTGATGGTCNG	359
Qy	123	ArgGlnArgSer-----IleSerSerGlyValAsnThrAlaIsthrProSerPro	139
Db	358	CGGAGAGCGGTGGGGCCGGCGGCTTCACCGGGGTACTCGCGGCAATGCGCAACACCT	299
Qy	140	ValAlaValGlnSerSerThrProProValGlnIleHisProAlaValGlnLysProThr	159
Db	298	TCTGTATACAGAGGCGTGGCCCCGGT-----CCCCAAGAGTCCGGCGGCG	248
Qy	160	ProProValValValLysLysProThrProThrProProProValGlnGlnProAla	179
Db	247	CCGCGCGTGG-----CCGCTGCCCGCGCCCGCTG-----CCGAGG	212
Qy	180	Pro-----ValAlaProProValThrGluAla	188

Db 211 CCGGCGGCAAGGGCCAGAGGGCCAGAGAAATGAGTGGCCCGACCGGTGTGTGC -CG 153
 |||
 |||
 Oy 189 Pro-----PheAlaIrnGlySerSerGlyValMetGlnPheArgGlyProValGly 205
 |||||
 |||||
 Db 152 CCGGCGGTCGCTGCGCGGTGGAGGCTTGGGTGTGCATCC----- 111
 |||||
 |||||
 Oy 206 AAlaIrnAsnProValValAlaArgArgPheGlyIrnAlaThrValAlaGlySerThrValIrn 225
 |||||
 |||||
 Db 110 -----GGTGGCGCTGTGTCCCGCCCGGCAATGACTCTTA 78
 |||||
 |||||
 Oy 226 SerAsnGlyMet-TripheSerGlyIaArgAspGlyIaSerPheIaAsnAlaSerAsnAlaGly 245
 |||||
 |||||
 Db 77 TGTAAAGGCGCTGTGTCTCTCAAAAAAGGGGGCCCGGTACNCCAAATGCCCTTAATAGG 18
 |||||
 |||||
 Oy 245 yrnValIleGlnAla 250
 |||||
 |||||
 Db 17 AGTCGTATTACATCC 2

	RESULT 2	DNA	906 bp	linear	GSS 12-JUN-2001
CNS01JER/C	Locus	CNS01JER	Anopheles gambiae GSS T7 end of clone 13F15 of Notridamel library from strain PEST of Anopheles gambiae (African malaria mosquito), genomic survey sequence.		
	DEFINITION				
	ACCESSION	AL146948			
	VERSION	ALI46948.1	GI:7005094		
	KEYWORDS	GSS.			
	SOURCE	African malaria mosquito.			
ORGANISM		Anopheles gambiae			
		Eukaryota; Metazoa; Arthropoda; Insecta; Peiryocta; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anophelinae.			
REFERENCE		1 (bases 1 to 906)			
AUTHORS	TITLE	Genoscope.			
JOURNAL		Direct Submission			
		Submitted (16-FEB-2000) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : sequefr@genoscope.cns.fr - Web : www.genoscope.cns.fr)			
REFERENCE		2 (bases 1 to 906)			
AUTHORS		Roth,C.W., Bray,P.T., Ke,Z., Collins,F.H. and Weissenbach,J.			
TITLE		Direct Submisson			
JOURNAL		Submitted (12-FEB-2000) BWHI, Institut Pasteur, 25, rue du Dr. Roux, Paris 75015, France			
COMMENT		This clone is from an A. gambiae BAC library provided by F.H. Collins and sequenced by A. gambiae in collaboration with the Laboratory of Biochem. and Biol. Molec. of Insects, Institut Pasteur.			

FEATURES	Location/Qualifiers
source	1..906
	/organism="Anopheles gambiae"
	/strain="PEST"
	/db_xref="taxon:7165"
	/clone="13P15"
	/clone_1b="Notredame1"
	/note="end : T7"
BASE COUNT	200 a 242 c 248 g 214 t 2 others
ORIGIN	
Alignment Scores:	
Pred. No.:	0.0329 Length: 906
Score:	136.00 Matches: 51
Percent Similarity:	36.74% Conservative: 28
Best Local Similarity:	23.72% Mismatch: 80
Query Match:	8.14% Indels: 57
DB:	17 Gaps: 4
US-10-018-706-2 (1-322) x CNS01JER (1-906)	
QY	105 TTTThrllelTrrThnglYgIntRpleuThreLleuTrpSerGlYasPleuYsValArgGlu 124 :: ::
Db	903 TACACATTCGA-GGTGGGACGACGACCTAAAGTT---AGCGGCGAGTCGCGCGGCGCAA 848

```

Qy 125 ArgSerIleSerSergIyValAsnThrAlaHisThrProSerProValAlaValGlnSer 144
Db 847 AAGTCTCTTGGAGAGGCGAAAACCGCCAAAGTACCGCTCTATACAGTCCGGAATCG 788
Qy 145 SerArgProProValGlnGlnHisProAlaValGlnIlyProThrProProValVal 164
Db 787 TCATGCGCCGCGGTGGACAGCGCTGCTGATTTGGCCCTGACGCGGTAAGTGCTC--- 731
Qy 165 ValIlyIysProThrProThrProThrProProValValGlnGlnProAlaProValAlaProPro 184
Db 731 ----- 731
Qy 185 ValThrGlnAlaProPheAlaThrGlySerSergIyValMetGlnPheArgTyProVal 204
Db 730 -----GCCCTTACTACTCTCTGAAAGTGC----- 704
Qy 205 GlyAlaThrAsnProValValAlaArgPheGlyThrAlaThrValAlaGlySerThrVal 224
Db 704 ----- 704
Qy 225 ThrSerAnGlyMetTrpPheSergIyArgAspGlyAspLeuIleAsnAlaSerAsnAla 244
Db 703 ---AACAGGGTATTGAGTATTTCGCTGCGCGGCTACGCGGTTTATGCTCCGGAAGCC 647
Qy 245 GlyThrValIleGlnAlaAspHisAsnMetAspGlyAlaSer-----IleValIleGln 262
Db 646 GGGAAAGGTGTTTACGTGCTGTAACGAGCTGCTGCTACGTAACCTGATCATGATTAG 587
Qy 263 HisThrAnGlyPheValSerSergTyrlleHisIleIyAspAlaGlnValIlySerThrGly 282
Db 586 CATGCGGAAGACTACATCATCAGCCCTATGCGCAACAAGCACAGATGCTGTAATTAACGGG 527
Qy 283 AspThrValAlaGlnThrGlyGlnArgIleAlaSerMetIyAsnGlnProSergIyAlaAla 302
Db 526 CAGAACTGTAAGCGCGGAGAAAGATTGCAACATGAGCGACGCGTAAGCATTCAGTG 467
Qy 303 LeuPheGlnPheArgIleSerArgAnGlyValTyValAspPro 317
Db 466 AAGCTGATTTCCAGATCCGCTATAAGCGACGCGCATCATCCG 422

RESULT 3
BH770991 4812 bp DNA linear GSS 01-MAY-2002
LOCUS LMGtag715 MG1363 Random Sequence Tag Library Lactococcus lactis
DEFINITION subsp. cremoris genomic, DNA sequence.
ACCESSION BH770991
VERSION BH770991.1 GI:20373948
KEYWORDS GSS.
SOURCE Lactococcus lactis subsp. cremoris.
ORGANISM Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Lactococcus.
REFERENCE 1 (bases 1 to 4812)
AUTHORS BojotIn,A., Ehrlich,S.D. and Sorokin,A.
TITLE Studies of genomes of dairy bacteria Lactococcus lactis
JOURNLS Sci. Aliments. (2002) In press
COMMENT Contact: Sorokin A
Geneticque Microbieme
INRA
CRJ INRA, Domaine de Villvert, 78352 Jouy en Josas cedex, France
Tel: 33 1 34 65 25 16
Fax: 33 1 34 65 25 21
Email: sorokine@jouy.inra.fr
best homologue in strain H1403 is ycgg (96%)
Class: shotgun
High quality sequence start: 30
High quality sequence stop: 4784.
location/Qualifiers
1..4812
/organism="Lactococcus lactis subsp. cremoris"
/strain="MG1363"
/db_xref="taxon:1359"
/clone_lib="MG1363 Random Sequence Tag Library"

```

```

/note="Vector: pSGM02; Site 1: SmaI; Library of
chromosomal fragments of L.lactis strain MG1363 was
prepared by partial AluI digestion or by sonication."
BASE COUNT 1468 a 975 c 800 g 1569 t
ORIGIN
Alignment Scores:
Pred. No.: 0.551 Length: 4812
Score: 134.00 Matches: 74
Percent Similarity: 37.37% Conservative: 37
Best Local Similarity: 24.92% Mismatch: 97
Query Match: 8.02% Indels: 89
DB: 17 Gaps: 13

US-10-018-706-2 (1-322) x BH770991 (1-4812)
Qy 26 ThrThrCysIleuAlaGlyCysAlaSerIyProThrTyAsnSerThr----- 42
Db 102 ACTGCTGCTTGGACGGGAGATATGATACGACCGACCAAGTATGCGCTTCACTGATATCGC 161
Qy 43 -----SergIySergIySerHisArgThrSergIy 52
Db 162 ATTATTTCTCAATTAATTAATTGACTGTTTGGACGGAGCTTCTTCACTGGAATATCTAT 221
Qy 53 SergIyGlyLeuAlaIleGlySergIyValIleThrAspSergIyValProAsn--- 71
Db 222 TCTGTGTGC-----TGACACACCAATTTACGAATTAATATTTGGAACCAATATGC 272
Qy 72 -----ArgTyGlnValIlyGlnGlyAspThrValSerIyIleAlaGlnArgTy 88
Db 273 AGTTCAACTACTATATACCGTCAATCTGATGATCTCTTGGGAAATCTCAAAAGATAT 332
Qy 89 GlyLeuAsnTrpArgGlnIleGlyHisIleAsnAsnLeuAsnSerTyrlleTy 108
Db 333 GGAATTATGTGCTGCTCAAAATTCAAAGTGCAGATTAATTTAAAGTACC---ATTATCTAC 389
Qy 109 ThrGlyGlnTrpLeuThrLeuTrpSergIyAspLeuIyValArgGluArgSerIleSer 128
Db 390 ATTGCTCAAAACTTGTACTGACAGGTCA-----GCTTCTTCTACAAATTC 437
Qy 129 SergIyValAsnThrAlaHisThrProSerProValAlaValGlnSerSergArgPro 148
Db 438 GGTGTTTCAAAACAATTC-----GCACACACTACTCCCAACC 473
Qy 149 ValGlnGlnHisProAlaValGlnIlyProThrProProValValIlyValIyProPro 168
Db 474 ACTTCTGTGACACCTGCT-----AAACCAACTTCACAAACAACCTGTAAGAGGAATCC 527
Qy 169 ThrProThrProProValValGlnGlnProAlaProValAlaProProValThrGlnAla 188
Db 528 GGAGATACCTTGTGGCGCTATCGTAATATTAACCTAGTA----- 569
Qy 189 ProPheAlaThrGlySerSergIyValMetGlnPheArgTyProValGly-AlaThrAs 208
Db 570 ---GTGCTCAATGAAGAATCGGAATCATTAAGTTACATCACTTATATTTGCTCAAAA 626
Qy 208 nProValValArgArgPheGlyThrAla-----ThrValAlaGlySerThrIy 224
Db 627 TCTTATTTTTCACAAATCTGCTGCTCAATCTTCAACAGCTTCACTCACTGC 686
Qy 224 lThrSerAnGlyMetTrpPheSergIyArgAspGlyAspLeuIleAsnAlaSerAsnAl 244
Db 687 TACCAATATAC-----TCAAACTC 704
Qy 244 aglyThrValIleGlnAlaAspHisAsnMetAspGlyAlaSerIleValIleGlnHis 264
Db 705 G-----AC 707
Qy 264 rAnGlyPheValSerSergTyrlleHisIleIyAspAlaGlnValIyThrGlyAspTh 284
Db 708 TTTCTTACTCAATATGCTCAATTCAT-----AAGCTGTTAAAGAGATATC 755
Qy 284 rVal-----ArgThrGlyGlnArgIleAlaSerMetIyS 295

```

Db 756 TCCTGTGGGACTTTCGCAAAATCTGGCAGCCCAATGCTTCATCAAG 804
RESULT 4
LOCUS BQ791034
DEFINITION BQ791034 433 bp mRNA linear EST 30-JUL-2002
E3832 Chinese cabbage etiolated seedling library Brassica rapa
subsp. pekinensis cDNA clone E3832, mRNA sequence.
ACCESSION BQ791034
VERSION BQ791034.1 GI:22005996
KEYWORDS EST
SOURCE Brassica rapa subsp. pekinensis.
ORGANISM Brassica rapa subsp. pekinensis
Pultrayota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 433)
AUTHORS Kyu,S.H., Yang,K.A., Lee,S.Y., Kim,H.-I., Cho,M.J. and Lim,C.O.
TITLE Expressed Sequence Tags of Chinese Cabbage Etiolated Seedling cDNA
(2002)
JOURNAL Unpublished (2002)
COMMENT Contact: Lim, C.O.
Plant Molecular Biology & Biotechnology Research Centre
Gyeongsang National University
#900 Gazwa-dong, Jinju 660-701, Korea
Tel: 82 55 751 6255
Fax: 82 55 759 9363
Email: colim@nongae.gsnu.ac.kr
Seq primer: T7.
FEATURES
source
1..433
/location/Qualifiers
/organism="Brassica rapa subsp. pekinensis"
/cultivar="Jangwon"
/db_xref="taxon:51351"
/clone="E3832"
/clone_1lb="Chinese cabbage etiolated seedling library"
/issue_type="Etiolated seedling"
/lab_host="XL-1 Blue"
/note="Vector: pSPORT 1; Site_1: Sal I; Site_2: Not I"
BASE COUNT 125 a 209 c 50 g 49 t
ORIGIN
Alignment Scores:
Pred. No.: 0.0263 Length: 433
Score: 132.00 Matches: 35
Percent Similarity: 41.86% Conservative: 1
Best Local Similarity: 40.70% Mismatches: 38
Query Match: 7.90% Indels: 12
DB: 14 Gaps: 2
US-10-018-706-2 (1-322) x BQ791034 (1-433)
Qy 128 SerSerGlyValAsnThrAlaHisThrProSerProValAlaValGlnSerSerArgPro 147
Db 87 AGTAGTGGCCACCGCCACCAACCAACCAACCCACCTGTGTAAAGCCACCAACACA 146
Qy 148 ProValGlnGlnHisProAlaValGlnLysProThrProProValAlaValAlaLysLys 167
Db 147 CCAACACCACTCCACAGTCGTAAAGCCACCAACCAACGCCACCTGTGTAAAGCCA 206
Qy 168 ProThrProThrProProValAlaGlnGlnProAlaProValAlaProValThr--- 186
Db 207 CCAACACCAACCAACCGGTGTAAAGCCACCAACCAACCCCTCTGTGTAAAGCACT 266
Qy 187 ---GluAlaProPheAlaThrGlySerSerGlyValMetGlnPheArgTyrProValGly 205
Db 267 CCAACACCAACCGGTGTAAAGCCACCAACCAACCCCTCTGTGTAAAGCACT 266
Qy 206 AlaThrAsnProValAla 211
Db 297 CCAACTCTCTCTGTGTAAAGCCACCAACCAACCCCTCTGTGTAAAGCACT 314
RESULT 5

AF094939/c
LOCUS AF094939 1096 bp DNA linear GSS 29-AUG-2000
DEFINITION AF094939 Salmonella typhimurium LT2, Lambda DASH II Salmonella
typhimurium genomic clone 226-T3, DNA sequence.
ACCESSION AF094939
VERSION AF094939.1 GI:4322781
KEYWORDS GSS.
SOURCE Salmonella typhimurium.
ORGANISM Salmonella typhimurium.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Salmonella
1 (bases 1 to 1096)
REFERENCE Wong,R.M.Y. and McClelland,M.
AUTHORS End Sequences of Salmonella typhimurium LT2 Lambda DASHII Clones,
TITLE Li-Cor
JOURNAL Unpublished (1999)
COMMENT Contact: McClelland M
Molecular Biology
Sidney Kimmel Cancer Center
3099 Science Park Road, San Diego, CA 92121, USA
Email: mclelland@lifesci.sdu.edu
Class: shotgun.
FEATURES
source
1..1096
/organism="Salmonella typhimurium"
/strain="LT2"
/db_xref="taxon:602"
/clone="226-T3"
/clone_1lb="Salmonella typhimurium LT2, Lambda DASH II"
/note="Vector: Lambda DASH II; sequenced using Li-Cor
sequencer"
BASE COUNT 239 a 323 c 272 g 262 t
ORIGIN
Alignment Scores:
Pred. No.: 0.0998 Length: 1096
Score: 132.00 Matches: 85
Percent Similarity: 34.06% Conservative: 40
Best Local Similarity: 23.16% Mismatches: 134
Query Match: 7.90% Indels: 108
DB: 17 Gaps: 14
US-10-018-706-2 (1-322) x AF094939 (1-1096)
Qy 42 ThrSerGlySerGlySerHisArgThrSerGlySerGlyGlyLeuAlaIleGlySerG 61
Db 1059 ACCGAGGCGCCAGCAACCAATCGATCAGCGCGGAGAGATGAAGCTATTCCTCAGCA 1000
Qy 61 nValIleThrAspSerGln-----GlyValProAsnArgTyrGlnValLysG 77
Db 999 TGAAGCTGACGATTAACCGCAGCGAGCAAGTCGGCGTC--CATGAATACGTGCTCCAC 943
Qy 77 nGlyAAPTThrValSerLysIleAlaGlnArgTyrGlyLeuAsnThrPAsGluIleGlyHis 97
Db 942 AGCGATAGCTTAAGCAGCATTCGTGATCAGTACGCAATGCATATGACGATATTAAGCG 883
Qy 97 s-----IleAsnAsnLeuAsnSerSerTyrThrIleTyrThrG 110
Db 882 ACTTGGCGCTTCGATTAAGAGAGCTGCGCAATGTGAANAATGGCCCAACGCTT----- 831
Qy 110 yGlnTPrPLeuThrLeuTPrSerGlyAspLeuLysValArgLysGserIleSerSerG 130
Db 830 -TCTGTGACACGTACCGCTGATGGCGATTTAAGCGGTGTGACATGGGAAGTGTCCGCG 772
Qy 130 yValAsnThrAlaHisThrProSer-ProVal-----AlaValGlnSerSer 146
Db 771 TGAACCGCTGATCTGATGCACTGCGCAACGGTTTAAATGACAGATGAAATGACACA 712
Qy 146 rGProPro-----ValGlnGlnHisProAlaVal 155
Db 711 GGGGAGACTGGGTCAACAGTCTGTAAGGTACGGTAGGGGTAGCTTGTGCGCAGCGC 652
Qy 155 aGlnLysProThrProProValAlaValLys-----LysProThrPro----- 170

```

Db      651 GAAAGAGCCCGTTTAAACGACGAGAAATGACGCGAGTGAATAAAGCCATGACAGTGCA 592
      170 -----
Qy      591 GATGATTTTGGCAAGCTGAAAGGCGAGTGAATTTTGGTTCTGATGTCGGCGAGAT 532
      171 -----
Qy      531 GCTGATGCAAGCGTGAACAGAGTCACTTTGTCGGCGTGGAGATGCTTCGATGTA 472
      176 TnglnProAlaProValAlaProProValThrGlnAlaProPheAlaThrGlySer-Ser 195
      471 AGATTACTACCCGTCATTCGCGCGTGAAGTAAATTATGATACCGTATGGGGTGGTCT 412
Qy      196 GlyValMetGlnPhe-ArgTyrProValGly----- 205
      411 GCGCAAGGCTTTTATTCGCTCCGACCGCTAAACAGTTCCGATCTCTCCAACTTCAA 352
Qy      206 -----AlaThrAsnProValAlaArgArgPheGlyThr-----Al 217
      351 TCCGCGTGTCTGAACCCGCTTACCGACGCGTTCGCGCATCGCGCTGACTTGC 292
Qy      217 aThrValAlaGlySerThrValThrSerAsnGlyMetTrpPheSerGlyArgAspGlyAs 237
      291 GATGCCCGACGGGTACGCGGTGCTGTGCGTGGG-----GATGGCGA 250
Qy      237 PheIleAsnAlaSerAsnAlaGlyThrValIleGlnAlaAspHisAsnMetAspGlyAl 257
      249 GGTGCGTGTCTGAACCGTAAACGACGCGCTGCC-----GGTTA 214
Qy      257 aSerIleValIleGlnHisThrAsnGlyPheValSerSerTyrIleHisIleValAspAl 277
      213 CTACATTCGCGATTTGCTATGACGACCACTACACACAGTTACATGCACTTCCGTAAGCT 154
Qy      277 aGlnValIleThrGlyAspThrValArgThrGlyGlnAlaIleAlaSerMetLysAsn-- 296
      153 GCTGTGTGAACCGGCGGCAAAAGTGAACGTCGATGTAATGCGCTTCTGTGAACAC 94
Qy      297 ----GlnProSerGlyAlaAlaLeuPheGlnPheArgIleSerArgAsnGlyValTyrVa 315
      93 CGGCGCTTCCACAGCGCGCATCTG--CAATTATGAGTATGATCAACCAAGCCGCT 37
Qy      315 lAspProLeuThr 319
      36 TAACCTCTGACA 24

RESULT 6
LOCUS   BG543924 596 bp mRNA linear EST 01-MAY-2002
DEFINITION El673 Chinese cabbage etiolated seedling library Brassica rapa
ACCESSION BG543924
VERSION  BG543924.1 GI:20374904
KEYWORDS EST.
SOURCE  Brassica rapa subsp. pekinensis.
        Bukarjota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
        Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
        Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 596)
AUTHORS  Ryu,S.H., Kang,J.S., Kang,C.-h., Kim,C.Y., Choi,Y.J., Lee,S.-H.,
        Bahk,J.D., Lee,S.Y., Cho,M.J. and Lim,C.O.
        Expressed Sequence Tags of Chinese Cabbage Etiolated Seedling cDNA
        Unpublished (2001)
        Contact: Lim, C.O.
        Plant Molecular Biology & Biotechnology Research Centre
        Gyeongsang National University
        #900 Gajwa-dong, Jinju 660-701, Korea
        Tel: 82 55 751 6255
        Fax: 82 55 759 9363
        Email: colim@nongae.gsnu.ac.kr
        Seq primer: T7.
  
```

```

FEATURES
source
location/Qualifiers
1..596
/organism="Brassica rapa subsp. pekinensis"
/cultivar="Jangwon"
/db_xref="taxon:51351"
/clone="E1673"
/clone_1db="Chinese cabbage etiolated seedling library"
/tissue_type="Etiolated seedling"
/lab_host="XL-1 Blue"
/note="vector: pSPORT 1; Site_1: Sal I; Site_2: Not I"
BASE COUNT 154 a 229 c 100 g 113 t
ORIGIN
Alignment Scores:
Pred. No.: 0.0512 Length: 596
Score: 131.00 Matches: 33
Percent Similarity: 43.59% Conserved: 1
Best Local Similarity: 42.31% Mismatch: 32
Query Match: 7.84% Indels: 12
DB: 12 Gaps: 2
US-10-018-706-2 (1-322) x BG543924 (1-596)
Qy      136 ThrProSerProValAlaValGlnSerSerArgProProValGlnGlnHisProAlaVal 155
      32 ACACCAACCCACCTGTTGTATGACGACCAACACCAACCAACCACTCCACGTCGTT 91
Db      156 GlnLysProThrProProValAlaValAlaValLysProThrProThrProProValAl 175
      92 ACGCCAACCAACACGACGCGCACCTGTTGTATGACGACCAACCAACCAACGCGTGTGA 151
Qy      176 GlnGlnProAlaProValAlaProProValThr-----GlnAlaProPheAlaThrGly 193
      152 ACGCCACGACACCAACCCCTCTGTCGTATGACCTTCAACACCAACGCGTGTACG--- 208
Qy      194 SerSerGlyValMetGlnPheArgTyrProValGlyAlaThrAsnProValAl 211
      209 -----CAACGACCAACCACTCTCTGTGTGA 235

RESULT 7
LOCUS   BM921488 1306 bp mRNA linear EST 12-MAR-2002
DEFINITION AGENCOURT 6626294 NIH_MGC_115 Homo sapiens cDNA clone IMAGE:5752955
ACCESSION BM921488
VERSION  BM921488.1 GI:19371867
KEYWORDS EST.
SOURCE  human.
ORGANISM Homo sapiens
        Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
        Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 1306)
AUTHORS  NIH-MGC http://mgs.nci.nih.gov/.
        National Institutes of Health, Mammalian Gene Collection (MGC)
        Unpublished (1999)
        Contact: Robert Strausberg, Ph.D.
        Email: cgabs-remail.nih.gov
        Tissue Procurement: Life Technologies, Inc.
        cDNA Library Preparation: Life Technologies, Inc.
        cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
        DNA Sequencing by: Agencourt Bioscience Corporation
        Clone distribution: MGC clone distribution information can be
        found through the I.M.A.G.E. Consortium/LNL at:
        http://image.llnl.gov
        Plate: LHAM2787 row: p column: 12
        High quality sequence start: 90
        High quality sequence stop: 471.
        Location/Qualifiers
1..1306
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5752955"
/clone_1db="NIH_MGC_115"
  
```

/lab host="DH10B"
 /note="Organ: pooled brain, lung, testis; Vector:
 pCMV-SPORT6; Site 1: NotI; Site 2: EcoRV (destroyed); RNA
 source anonymous pool of 6 male brains, age range 23-27; 1
 male lung, age 27; and 1 male testis, age 69. Library is
 oligo-dT primed and directionally cloned (EcoRV site is
 destroyed upon cloning). Average insert size 1.8 kb,
 insert size range 1-3 kb. Library is normalized and
 enriched for full-length clones and was constructed by C.
 Gruber (Invitrogen). Research Genetics tracking code
 021. Note: this is a NIH-MGC Library."

BASE COUNT 223 a 524 c 344 g 213 t 2 others

ALIGNMENT SCORES:

Pred. No.: 0.158 Length: 1306
 Score: 131.00 Matches: 54
 Percent Similarity: 39.60% Conservative: 26
 Best Local Similarity: 26.73% Mismatches: 91
 Query Match: 7.84% Indels: 31
 DB: 14 Gaps: 9

US-10-018-706-2 (1-322) x BM921488 (1-1306)

QY 42 ThSerGlySerGlySerHisArgThrSerGlyGlyValLeuAla---IleGlySer 60
 DB 147 ACCACCACTCTGGGGTTCAGAGAAATGAAAGCTTGGGGTCTTAAATTGGAGCT 206
 QY 61 Gln-----ValIleThrAspSerGlnGlyValProAsnArgTyr 73
 DB 207 GAATGGCCCATGTATGACGGTCCCTGATGAGGCTCTTGGCCCTTGAATGCTA 266
 QY 74 Gln-----VallyGlnGlyAspThrValSerIleLeuIleArgTyrGly 89
 DB 267 CTCACATCGACATTCAGAGACTGAGAGTTAGGCTTCAAGCATCCAGCTTCTCC 326
 QY 90 LeuAsnTrpArg-----GluIleGlyHisIleAsnAsnLeuAsnSerSerTyr 105
 DB 327 TTTCATTGGAACCTTGACATTTTCCGCCAGGCTTTTGGAAAGCCAAATGAGCTCAG 386
 QY 106 ---ThrIleTyrThrGlyIntTrpLeuThrLeuTrpSerGlyAspLeuys----- 121
 DB 387 AAGCTGTATGACCTCTGTGTGGTATATCTTGTCTGCGATGTTTATTTATTTTC 446
 QY 122 ---ValArgGlnArgSerIleSer-SerGlyValAsnThrAlaHisThrProSerProVa 140
 DB 447 TGAAGCCAGACATGATCAGACAGGTGCGGTGACAACTTCAGCACTTACACCATAT 506
 QY 140 AlaValGlnSerSerArgProProValGlnGlnHisProAlaValGlnGlnProThrPr 160
 DB 507 GGCT-----TCGAGGCCCGCGGTTCCCTCAGCGAGCCGAGCGAGCCGACTGCC 557
 QY 160 oProValValValValysbProThrProThrProProValValGlnGlnProAlaPr 180
 DB 558 ACCCTAC-----CCCTTACCCCGCCGCCCAACCCATACCCCGCCGAGGAGACCG 605
 QY 180 oValAlaProProValThrGlnAlaProPheAlaThrGlySerSerGlyValMecGlnPh 200
 DB 606 CGCCCCCCCCCGCCCCCGCCCCCGAGCCACCCCGCCCCCGCCGCTGGAGTCA 665
 QY 200 eArgTyrProVal-----GlyAlaThrAsnProValAlaArgArgPheGlyThrAlaTh 218
 DB 666 CGCGCCCCCACTCCCGCCCGCGAGCGCGCGCCGAGCCCGCCCGCTGGAGCGGACATC 725
 QY 218 rVal 219
 DB 726 CGTC 729

RESULT 8
 BQ791640 398 bp mRNA linear EST 30-JUL-2002
 LOCUS BQ791640
 DEFINITION E3323 Chinese cabbage etiolated seedling library Brassica rapa
 subsp. pekinensis cDNA clone E3323, mRNA sequence.

ACCESSION BQ791640 GI:22006602
 VERSION BQ791640.1
 KEYWORDS EST.
 SOURCE Brassica rapa subsp. pekinensis.
 ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Brassicaceae; Brassicaceae; Brassicaceae; Brassicaceae; Brassicaceae;
 Rosidae; eudicotids II; Brassicales; Brassicaceae; Brassica.

REFERENCE 1 (bases 1 to 398)
 AUTHORS Ryu, S.H., Yang, K.A., Lee, S.Y., Kim, H.-I., Cho, M.J. and Lim, C.O.
 TITLE Expressed Sequence Tags of Chinese Cabbage Etiolated Seedling cDNA
 JOURNAL Unpublished (2002)

COMMENT Contact: Lim, C.O.
 Plant Molecular Biology & Biotechnology Research Centre
 Gyeongsang National University
 #900 Gajwa-dong, Jinju 660-701, Korea
 Tel: 82 55 751 6255
 Fax: 82 55 759 9363
 Email: colim@nongae.gsnu.ac.kr
 Seq primer: T7.

FEATURES
 source
 1..398
 /organism="Brassica rapa subsp. pekinensis"
 /cultivar="Jangwon"
 /db_xref="taxon:5151"
 /clone="E3323"
 /clone_lib="Chinese cabbage etiolated seedling library"
 /tissue_type="Etiolated seedling"
 /lab_host="XL-1 Blue"
 /note="Vector: pSPORT 1; Site_1: Sal I; Site_2: Not I"

BASE COUNT 105 a 165 c 67 g 61 t

ALIGNMENT SCORES:

Pred. No.: 0.0392 Length: 398
 Score: 129.50 Matches: 33
 Percent Similarity: 48.57% Conservative: 25
 Best Local Similarity: 47.14% Mismatches: 11
 Query Match: 7.75% Indels: 11
 DB: 14 Gaps: 3

US-10-018-706-2 (1-322) x BQ791640 (1-398)

QY 131 ValAsnThrAlaHisThrProSerProValAlaValGlnSerSerArgProProValGln 150
 DB 79 GTGCTAAGCGCACCGACCAACCCCTCTGTGTAACACCTTCAACACACCGGTCGTA 138
 QY 151 GlnHisProAlaValGlnysProThrProProValVal-----ValI 164
 DB 139 ACCGACACCGACA-----CCACTCTCTCTGTGTGTAACACCAACCAACCGGTC 189
 QY 165 VallysbysProThrProThrProProValValGlnGlnProAlaPro-----ValAla 182
 DB 190 GTAAGCGCACCGACCAACCCCTCTGTGTGTAACACCTTCAACACCAACCGGTCGTA 249
 QY 183 ProProValThrGlnAlaProPheAlaThr 192
 DB 250 CCACCGACCAACCCCTCTGTGTGTAAC 279

RESULT 9
 CNS03D9V/c 890 bp DNA linear GSS 15-MAY-2000
 LOCUS CNS03D9V/c
 DEFINITION Tetradodon nigroviridis genome survey sequence 17 end of clone
 019021 of library G from Tetradodon nigroviridis, genomic survey
 sequence.

ACCESSION AL239728 GI:7898863
 VERSION AL239728.1
 KEYWORDS GSS; genome survey sequence.
 SOURCE Tetradodon nigroviridis.
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
SOURCE

Acanthomorphs; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.
1 (bases 1 to 890)
Roest-Crollius,H., Jallion,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brotier,P., Quetier,F., Saurin,W. and Weissenbach,J.
Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence
Unpublished
2 (bases 1 to 890)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.
Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis
Unpublished
3 (bases 1 to 890)
Genoscope.
Direct Submission
Submitted (12-APR-2000)
This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at <http://www.genoscope.cns.fr/tetraodon>.
Location/Qualifiers
1..890
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="019021"
/clone_1lb="G"
/note="Genoscope sequence ID : COBG019AH11P1-end : T7"
BASE COUNT 172 a 243 c 338 g 128 t 9 others
ORIGIN

Alignment Scores:
Pred. No.: 0.21 Length: 890
Score: 127.00 Matches: 60
Percent Similarity: 36.73% Conservative: 23
Best Local Similarity: 26.55% Mismatches: 83
Query Match: 7.60% Indels: 60
DB: 17 Gaps: 9

US-10-018-706-2 (1-322) x CNS03DYV (1-890)

25 ILeThrThrCysIleLeuAlaGlyCysAlaSerIysProThrTyranSerThrSergly 44
Db 765 TTGACCTCTGCGACGCTCTCGGCTGCGAGCGGAGCTGCGACATCTGAGCGCGCTCC 706
Qy 45 Ser-----GlySerH1sArgThrSerglySerglyLeuAlaIleGlySerglyVal 62
Db 705 GAGTCTTTTGAGMAAAGCGCACCGTGGAGC-----ATG 670
Qy 63 ILeThrAspSerGlnGlyValProAsnArgTyrGlnVallyGlnGlyAspThr----- 80
Db 669 TACCGCAGCGCGCGGTACAGTTGTTTGGAAATTAACCTGAGTACACCCCGTTT 610
Qy 81 -----ValSerIysIleAlaGlnArgTyrGlyLeuAsnTrpArgGlnIleGlyH1sIle 98
Db 609 CACCTTACGAGAGAAACACGAGAGAGTGGGCTCAAC----- 571
Qy 99 AsnAsnLeuAsnSerIysTyrH1sIleTyrThrGlyGlnTrpLeuThrLeuTrpSergly 118
Db 570 -----AGCTGACCTGTTTGAACAAGCGTAACCTCAACATGCTGG----- 532
Qy 119 AspleuLysValArgGluArgSerIleSerSerglyValAsnThrAlaH1sThrProSer 138
Db 531 -----TTTCAGAGGTAAACCCCTCCCGCTTGGGGGCTCT 499
Qy 139 ProValAlaValGlnSerSerArgProProValGlnGlnH1sProAlaVal----- 155
Db 498 CCGTGGGGCTGTCTGAGTCCCGACAGTCCCGACGTCTCCCGACAGTCTCCCGACGT 439
Qy 156 -----GlnLysProThrProProValValValValys 167

Db 438 CCCCCCAGCCCCCAGTCTCTCCAGTCCCGAGTCCCGCCAGTCCCGAGTCTCTCC 379
Qy 168 ProThr-----ProThrProProValValGlnGlnProAlaPro-ValAlaProProValTh 186
Db 378 CCGATCCCCCCCCAGTCTCTCCAGTCCCGAGTCTCTCCCGACAGTCTCT 325
Qy 186 rGluAlaProPheAlaThrGlySerSerglyValMetGlnPheArgTyrProValGlyAl 206
Db 324 CCGAGTCCCGCCAGTCCCGCCAGTCTCT-----CCAGTCCCGCA 286
Qy 206 arhAsnProValValArgArgPheGlyThrAla-ThrValAlaGlySerThrValThrs 226
Db 285 GTCTTCCCGAGTCCCGCCAGTCCCGCCAGTCTCTCAACAGTGGAGCGGCTTAACCCGCCA 226
Qy 226 eArgnGlyMetTrp 230
Db 225 GCAGTGGTCTGTGG 212

RESULT 10
B0704548
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
SOURCE

B0704548
Bn01_03006_A
Bn01_AAFc_ECORC transgenic Brassica napus overexpressing BNCBP17, co
nstitutively frost_tolerant Brassica napus cDNA clone Bn01_03006,
mRNA sequence.
B0704548
B0704548.1 GI:21843967
EST.
rape.
Brassica napus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
1 (bases 1 to 669)
Singh,J., Allard,G., Tinker,N., Robert,L., Lacroix,C., De Moors,A.,
Chagnon,J., Farah,S., Couroux,P. and Hattori,J.
Expressed Sequence Tags from constitutively frost tolerant
transgenic Brassica napus overexpressing BNCBP17
Unpublished (2002)
Contact: Singh,J.A.
Eastern Cereal and Oilseed Research Centre
Agriculture and Agri-Food Canada
KW Neaby Bldg., Central Experimental Farm, Ottawa, Ontario, K1A
0C6, Canada
Tel: (613) 759-1662
Fax: (613) 759-1701
Email: singhja@em.agr.ca.
Location/Qualifiers
1..669
/organism="Brassica napus"
/cultivar="Westar"
/db_xref="taxon:3708"
/clone="Bn01_03006"
/clone_1lb="Bn01_AAFc_ECORC transgenic Brassica napus over
expressing BNCBP17 constitutively_frost_tolerant"
/tissue_type="fourth leaf"
/dev_stage="3 weeks seedling grown at room temperature"
/note="Vector: Bluescript SK+/XhoI-EcoRI; Site 1: EcoRI;
Site 2: XhoI; Germinated in soil flats and seedlings grown
for 3 weeks in a Conviron E-15 cabinet set at 20°C/16 hr
light (250 Em-2sec-1) and 16°C/8 hr dark. Fourth leaves
collected at 9 am and immediately frozen."
BASE COUNT 167 a 203 c 124 g 168 t
ORIGIN

Alignment Scores:
Pred. No.: 0.155 Length: 669
Score: 126.50 Matches: 32
Percent Similarity: 48.57% Conservative: 2
Best Local Similarity: 45.71% Mismatches: 25
Query Match: 7.57% Indels: 11
DB: 14 Gaps: 3

US-10-018-706-2 (1-322) x BQ704548 (1-669)

QY 131 ValAsnThrAlaHisThrProSerProValAlaValAlaGlnSerSerArgProProValGln 150
 DB 35 GTCTGTAACCGCCACCAACCAACCTCTGTCGTAAACCAACCAACCAACCGGTCTGTA 94
 QY 151 GlnHisProAlaValGlnHisProProProValAla-----Val 164
 DB 95 AGCCGACCGACA-----CCAACTCTCTGTCGTAAACCAACCAACCAACCGGTCTG 145
 QY 165 ValValLysProThrProThrProProValAlaGlnGlnProAlaPro-----ValAla 182
 DB 146 ATACGCCACCGCCACCAACCAACCTCTGTCGTAAATCTCTCAACCAACCGGTCTGTAACG 205
 QY 183 ProProValThrGlnAlaProPheAlaThr 192
 DB 206 CCACCGACACCAACCCCTCTGTCGTAAAC 235

RESULT 11
 AVS20024 628 bp mRNA linear EST 06-SEP-2000
 LOCUS AVS20024 Arabidopsis thaliana aboveground organs two to six-week
 DEFINITION old Arabidopsis thaliana cDNA clone AP204g09F 3', mRNA sequence.
 ACCESSION AVS20024
 VERSION AVS20024.1 GI:8679551
 KEYWORDS EST.
 SOURCE thale cress.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; Eurosid II; Brassicales; Brassicaceae; Arabidopsie.
 1 (bases 1 to 628)
 REFERENCE Asamizu E., Nakamura Y., Sato S. and Tabata S.
 A large scale analysis of cDNA in Arabidopsis thaliana: Generation
 of 12,028 non-redundant expressed sequence tags from normalized and
 size-selected cDNA libraries
 DNA Res. 7, 175-180 (2000)
 JOURNAL 20363093
 MEDLINE
 COMMENT Contact: Erika Asamizu
 The First Laboratory For Plant Gene Research
 Kazusa DNA Research Institute
 Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
 Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/
 location/Qualifiers

FEATURES
 source
 1..628
 /organism="Arabidopsis thaliana"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /clone="AP204g09F"
 /clone_1b="Arabidopsis thaliana aboveground organs two to
 six-week old"
 /tissue_type="aboveground organs"
 /dev_stage="two to six-week old"
 /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
 XhoI"

BASE COUNT 157 a 105 c 197 g 169 t
 ORIGIN

Alignment Scores:
 Pred. No.: 0.174 Length: 628
 Score: 125.50 Matches: 31
 Percent Similarity: 46.25% Conservative: 6
 Best Local Similarity: 38.75% Mismatches: 32
 Query Match: 7.51% Indels: 11
 DB: 10 Gaps: 3

US-10-018-706-2 (1-322) x AVS20024 (1-628)

QY 131 ValAsnThrAlaHisThrProSerProValAlaValAlaGlnSerSerArgProProValGln 150
 DB 609 GTCTGTAACCGCCACCAACCAACCTCTGTCGTAAACCAACCAACCGGTCTGTAACG 559

QY 151 GlnHisProAlaValGlnLysProThrProProValValValLysLysProThrPro 170
 DB 558 ACTCCACCGGTATTAACCAACCAACCAACCTCTGTCGTAAACCAACCAACCGGTCTGTA 499
 QY 171 ThrProProValValGlnGlnProAlaProValAlaProProValThrGlnAlaProPhe 190
 DB 498 ACACCAACCGGTGTGACCTCAACACCA--ACACCAACCGGTGTGTAACCAACCAACA 442
 QY 191 AlaThrGlySerSerGlyValMetGlnPheArgTyProValGlyAlaThrAspProVal 210
 DB 441 CCNACTCTCTGTAACG-----CCTATTCCAGAACTTGCCCAATT 403

RESULT 12
 BM911051 1279 bp mRNA linear EST 12-MAR-2002
 LOCUS BM911051
 DEFINITION AGENCOURT 6613368 NIH_MGC_98 Homo sapiens cDNA IMAGE:5455034
 5', mRNA sequence.
 ACCESSION BM911051
 VERSION BM911051.1 GI:19361430
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 1279)
 REFERENCE NIH-MGC http://mgi.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 JOURNAL
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabs-remail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.llnl.gov
 Plate: LHCN951 row: c column: 03
 High quality sequence start: 6
 High quality sequence stop: 259.
 Location/Qualifiers

FEATURES
 source
 1..1279
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5455034"
 /clone_1b="NIH_MGC_98"
 /tissue_type="astrocytoma grade IV, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
 EcoRI; cDNA made by oligo-dT priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5'
 adaptor: GGCAAGAG(G). Library constructed by Ling Hong
 in the laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies).
 Note: this is a NIH_MGC Library."

BASE COUNT 332 a 531 c 231 g 184 t 1 others
 ORIGIN

Alignment Scores:
 Pred. No.: 0.485 Length: 1279
 Score: 125.50 Matches: 51
 Percent Similarity: 36.83% Conservative: 22
 Best Local Similarity: 27.13% Mismatches: 66
 Query Match: 7.51% Indels: 50
 DB: 14 Gaps: 7

US-10-018-706-2 (1-322) x BM911051 (1-1279)

QY 41 SerThrSerGlySerGlySerHisArgThrSerGlySerGlyValLeuAlaIleGlySer 60
 DB 100 TCTCTCGGCTGTCAGATCCCAACCGCTCGG-GGCCATGAGAGT--GCTTCCGACCGG 155

Qy 61 GlnValIleThrAspSerGlnGlyValProAsnArgTyrGlnVal-----Lys 76
 Db 156 AATGTACTACTGGAGTCTGGAGAGCGGAGTGGTGAAGGATGACACCGGGGGGCGC 215
 Qy 77 GlnGlyAspThrValSerIleValGlnArgTyrGlyLeuAsnTyrArgGlnIleGly 96
 Db 216 CTTGAAATCATGCGCGCTTGTCTTCGAGCGGGGAGACACCGGGGGGAGGGA 275
 Qy 97 HisIle-----AsnAsnLeuAsnSer 103
 Db 276 GGACTCTCCCAACAAATGCTCGCGCTCCCAATCCGGGTACGAATTCATGGAACCGA 335
 Qy 104 SerTyrThrIleTyrGlnGlnTyrLeuThrIleTyrPserGlyAspLeuValArg 123
 Db 336 ACCCACCACCATCATACGACCAATTCCTTAACGTTA----- 371
 Qy 124 GluArgSerIleSerSerGlyValAsnThrAlaHisThrProSerProValAlaValGln 143
 Db 372 -----CCACAAACCCACACACACCCCGACGACACCCCGCCCGCCACTACCA 422
 Qy 144 SerSerArgProProValGlnGlnHisProAlaValGlnGlnProThrProProValVal 163
 Db 423 TACCCCACTCTCCCTC-----CACCATCTCTCATCACCCCGCCACACT----- 470
 Qy 164 ValValIleValysProThrProThrProProValValGlnGlnProAlaProValAlaPro 183
 Db 471 -----AAACCCAGCGCCCGCCACACACCAAAAGAGCGCGACCGCAACCGAACCC 521
 Qy 184 ProValThrGlnAlaProPheAlaThrGlySerSerGlyValMetGlnPheArgTyrPro 203
 Db 522 GCCGTAAAGCAGCAGGACT-----CCT 542
 Qy 204 ValGlyAlaThrAsnProValVal 211
 Db 543 ACGGCGAGCATTAACCGCATCATC 566

RESULT 13
 BG593928 666 bp mRNA linear EST 12-APR-2001
 LOCUS BG593928
 DEFINITION EST12957606 cSTS Solanum tuberosum cDNA clone cSTS516 5' sequence,
 mRNA sequence.
 ACCESSION BG593928
 VERSION BG593928.1 GI:13612068
 KEYWORDS EST.
 SOURCE Solanum tuberosum
 ORGANISM Solanum tuberosum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
 1 (bases 1 to 666)
 van der Hoeven, R., Bezzerides, J., Sun, H., Cho, J., Chiemingo, A.,
 Bougri, O., Buell, C. R., Romning, C., Tankley, S. and Baker, B.
 Generations of ESTs from sprouting potato eyes
 Unpublished (2000)
 TITLE CONTACT: Cathy Romning
 JOURNAL The Institute for Genomic Research
 COMMENT For clone info: please contact Research Genetics, Libraries
 Division tel 1-800-711-6195, email cdna@resgen.com
 Seq primer: M13F-R.
 FEATURES
 SOURCE Location/Qualifiers
 1..666
 /organism="Solanum tuberosum"
 /cultivar="Kennebec"
 /db_xref="taxon:4113"
 /clone="cSTS516"
 /clone_id="cSTS"
 /tissue_type="sprouting eyes from tubers"
 /dev_stage="12-14 weeks post harvest"
 /lab_host="SOLR"
 /note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2:
 XhoI; Various sizes of sprouting eyes (2mm to 15mm) were
 taken from tubers. The tubers were incubated at 26C in the
 dark for 2-3 weeks prior to sprouting. The eyes were

BASE COUNT 139 a 212 c 134 g 181 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 0.234
 Score: 124.50
 Percent Similarity: 35.96%
 Best local Similarity: 27.53%
 Query Match: 7.46%
 DB: 12
 Gaps: 6
 US-10-018-706-2 (1-322) x BG593928 (1-666)
 frozen in liquid nitrogen immediately upon removal from
 tubers."

Qy 115 LeuThrSerGlyAspLeuValArgGluArgSerIleSerSerGlyValAsn----- 132
 Db 80 CTTGGATCGGGTTTCTGTGATT-----ATCGTGGTGGTGAAGGGGTCAA 127
 Qy 133 -----ThrAlaHisThrProSerPro 139
 Db 128 GCTCTGCAAGTCAACCCACCACTAGTCCCGACACCAACATCTACCTCCGACCT 187
 Qy 140 Val-----AlaValGlnSerSerArgProProValGlnGlnHisProAla 154
 Db 188 ACCGGTTCCTCACCTCTCGAGTGAATTCACCTCTCATCTGTATCTCTCCACCGGCA 247
 Qy 155 ValGlnValProThrProProValValValValValysProThrProThrProProVal 174
 Db 248 ACTTCAACCACTAGTACTCTTCCACGAGCGCTCCACCACTAGTACTTCCACCA 307
 Qy 175 ValGlnGlnProAlaProProValAlaProProValThrGlnAlaProPheAlaThrGlySer 194
 Db 308 GCTACTCTCTCCACAGAGTGTCTCTTCCACCTGTGATCTCAAC-TCACCGCGCAGC 366
 Qy 195 SerGlyValMetGlnPheArgTyrProValGlyAlaThrAsnProValAlaArgArgPhe 214
 Db 366 ----- 366
 Qy 215 GlyThrAlaThrValAlaGlySerThrValThrSerAsnGlyMetThrPheSerGlyArg 234
 Db 367 GGCACCTGACCGGTCGCAACTCCACTGCTTCGCTCC-----TGCTCTCTCTCGACTAC 423
 Qy 235 AspGlyAspLeuIleAsnAlaSerAsnAlaGlyThrValIleGlnAlaPheIleAsnMet 254
 Db 424 AAGGTAGCTACATCTCCGACCTTCCCGGAGGATGTGATGCTCTCCAGACACACC 483
 Qy 255 AspGlyAlaSerIleValIleGlnHisThrAsnGlyPheValSerSerTyrIle 272
 Db 484 TCTGGTGTCTCCGTC-----ACCGAGTGTATTATCTCTCTCATC 525

RESULT 14
 BQ505542 767 bp mRNA linear EST 22-JUN-2002
 LOCUS BQ505542
 DEFINITION EST612957 Generation of a set of potato cDNA clones for microarray
 analyses mixed potato tissues Solanum tuberosum cDNA clone STM6E76
 5' end, mRNA sequence.
 ACCESSION BQ505542
 VERSION BQ505542.2 GI:21921472
 KEYWORDS EST.
 SOURCE Solanum tuberosum
 ORGANISM Solanum tuberosum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
 1 (bases 1 to 767)
 Buell, C. R., Hart, A., Baker, B., Tankley, S., Fry, W., Smart, C.,
 Restrepo, S., Griffiths, H., van der Hoeven, R., Teal, J. and
 Karamycheva, S. A.
 Generation of a set of potato cDNA clones for microarray analyses
 Unpublished (2002)
 TITLE CONTACT: Cathy Romning
 JOURNAL The Institute for Genomic Research
 COMMENT On Jun 10, 2002 this sequence version replaced gi:21364411.
 Other ESTs: EST612958

Contact: Robin Buell
The Institute for Genomic Research
9712 Medical Center Dr, Rockville, MD 20850, USA
Email: potatob@tigr.org
This clone is available through the Research Genetics, contact the
Research Genetics for further information 1-800-711-6195 or
cdna@resgen.com
Seq primer: 13

FEATURES
source location/Qualifiers

1..767
/organism="Solanum tuberosum"
/cultivar="Kennebec or Bintje"
/db_xref="taxon:4113"
/clone_lib="Generation of a set of potato cDNA clones for
microarray analyses mixed potato tissues"
/issue_type="mixed tissues"
/lab_host="SOLR"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; supplier: Combination of untreated and Phytophthora
infestans-treated libraries of stolons, leaves, leaflets,
axillary buds of stem explants, petioles, germinating eyes
, tubers, or roots."

BASE COUNT 164 a 225 c 152 g 226 t
ORIGIN

Alignment Scores:

Pred. No.: 0.287 Length: 767
Score: 124.50 Matches: 49
Percent Similarity: 35.96% Conservative: 15
Best Local Similarity: 27.53% Mismatches: 65
Query Match: 7.46% Indels: 50
Gaps: 6

US-10-018-706-2 (1-322) x BQ505542 (1-767)

QY 115 leuTpsrGlyAspLeuLysValArgGluArgSerLysSerGlyValAsn----- 132
DB 80 CTTTGATCGGGGTTTCTGCGATT-----ATGTCGCTGGTGTAGGGGGTCA 127
QY 133 -----ThAlHisThrProSerPro 139
DB 128 GCTCTGCGACGTCACCACTAGTCCCGACACCAACTCACTACCTCCCTCACCCT 187
QY 140 Val-----AlaValGlnSerSerArgProProValGlnGlnHisProAla 154
DB 188 ACCGTTCTCCACCTCTCGACAGTCACTCCCTCCATCTGTATCTTCTCCACCGGCA 247
QY 155 ValGlnLysProThrProProValValValLysLysProThrProThrProProVal 174
DB 248 ACTTCAACCACTCTCTCTCCACAGAGCGCTCCACAGTGAAGTCTCCACACCA 307
QY 175 ValGlnGlnProAlaProValAlaProProValThrGlnAlaProPhenAlaThrGlySer 194
DB 308 GCTACCTCTCCACAGAGTGTCTCTCCACCTGTAGTACCTCACCTCCACCGGACG 366
QY 195 SerGlyValMetGlnHeArgTyProValGlyAlaThrAsnProValAlaArgArgPhe 214
DB 366 ----- 366
QY 215 GlyThrAlaThrValAlaGlySerThrValThrSerAsnGlyMetTrpPheSerGlyArg 234
DB 367 GGCACCTGGACCGGTGGCAACTGCACTGTTGAGCTCC--TGCTCTACTCCAGTAC 423
QY 235 AspGlyAspLeuLeuAsnAlaSerAsnAlaGlyThrValIleGlnAlaAspHisAsnMet 254
DB 424 AAAGGAGTACATCTCTGGCAGCTTGGCGGAGGAGGATGTAGTCCACGACCAACC 483
QY 255 AspGlyAlaSerIleValIleGlnHisThrAsnGlyPheValSerSerTrpIle 272
DB 484 TCTGGTGTCTCGTC-----ACGAGTGTATTATCTCTGCTCCATC 525

BF859163
LOCUS BF859163 839 bp mRNA linear EST 19-JAN-2001
DEFINITION 963001C10.Y3 C. reinhardtii CC-1690, Stress condition I, normalized
, Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION BF859163
VERSION BF859163.1 GI:12249254
KEYWORDS EST.
SOURCE Chlamydomonas reinhardtii.
ORGANISM Chlamydomonas reinhardtii
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
Chlamydomonadaceae; Chlamydomonas.
REFERENCE 1 (bases 1 to 839)
Grossman,A., Davies,J., Federpiel,N., Harris,E., Hauser,C.,
Lefebvre,P., McDermott,J.P., Shrago,J., Sillflow,C. and Stern,D.
Analyses of the Chlamydomonas reinhardtii Genome: A Model,
Unicellular System for Analyzing Gene Function and Regulation in
Vascular Plants; Project phase 3
JOURNAL Unpublished (2000)
COMMENT Contact: Charles Hauser
DCMB Box 91000
Duke University
Durham, NC 27708-1000
Tel: 919 613 8159
Fax: 919 613 8177
Email: chauser@duke.edu.
Location/Qualifiers
1..839
/organism="Chlamydomonas reinhardtii"
/strain="CC-1690 wild type mt+ 21gr"
/db_xref="taxon:3055"
/clone_lib="C. reinhardtii CC-1690, Stress condition I,
normalized, Lambda zap II"
/note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2:
XhoI; this library, constructed by John Davies and Jeffrey
McDermott, combines cDNAs from CC-1690 cells grown to
mid-log phase in TAP-N (30 min, 1hr, 4hr), TAP-S (30 min,
1hr, 4hr), TAP-P (4hr, 12hr, 24hr), NO3 to NH4 (30min, 1hr,
4hr) and NH4 to NO3 (30min, 1hr, 4hr). PolyA mRNA was
purified from each sample, pooled and cDNA synthesized.
The cDNA was directionally cloned into lambda zap II
(Stratagene) in the EcoRI (5') and XhoI (3') sites.
pBluescript II SK- plasmids were excised from the lambda
ZAP clones by superinfection with Exsist (Stratagene)
phage. The library was normalized using method 4 described
in Bonaldi et al (1996) genome Research 6: 791-806."
BASE COUNT 215 a 268 c 241 g 114 t
ORIGIN
Alignment Scores:
Pred. No.: 0.326 Length: 839
Score: 124.50 Matches: 41
Percent Similarity: 32.05% Conservative: 9
Best Local Similarity: 26.28% Mismatches: 53
Query Match: 7.46% Indels: 53
Gaps: 4
US-10-018-706-2 (1-322) x BF859163 (1-839)
QY 44 GlySerGlySerHisArgThrSerGlySerGlyLysLeuAlaIleGlySerGlnValIle 63
DB 67 GGTGGGGCGCCCTGAAGTGAAGTATGATCCCCGGGGCTG----- 105
QY 64 ThrAspSerGlnGlyAlaProAsnArgTyGlnValGlnGlyAspThrValSerLys 83
DB 106 -----CAGAAATTCGGACAGGACACCTAATGACAACTAGC----- 144
QY 84 IleAlaGlnArgTyGlyLeuEntPrpArgGluIleGlyHisIleAsnAsnLeuAsnSer 103
DB 144 ----- 144
QY 104 SerTyThrIleTyThrGlyGlnTrpLeuTrpLeuTrpSerGlyAspLeuLysValArg 123
DB 145 -----GCGACAGATGGGTTCTTGCTGGCAAGACCAAGCTGAAGTGGCG 189

Oy		124	GluAYserSeriLeSerSeGlyValaInhrxIahIstHrProProVaAlaValaGln	143
Dd		190	AACMAAAAAGCTAGTCCGTTCCTCGCCTCCGCCGCACCTGCTCCACACCAAGTGACA	249
Oy		144	SerSerArgProProVaL-----GlnGlnHisPro	153
Dd		250	CAGGACAGACGGCCCTGCTCCGAAGACTCGGGCGAAGTCAAGCCAGGCGCAGAACGAACT	309
Oy		154	AlaValGlnIlySProThProProVaIvalValValIlySLySProThProThProPro	173
Dd		310	CCTGGCCCCCGCCATCCCCACCG-----CCCACAGCTTAAGCTGCTCCTCT	357
Oy		174	ValValGlnGlnProAlaProAlaProProVaLProProVaLThrGlnAlaPro	189
Dd		358	GCGACGCCCCGGCGCCCTCCGGCTCCAGAAGTGTAGAGGCTATTCCCG	405
RESULT 16				
BH705150/c				
LOCUS				
DEFINITION				
ACCESSION				
VERSION				
KEYWORDS				
SOURCE				
ORGANISM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
COMMENT				
FEATURES				
source				
BASE COUNT				
ORIGIN				
Alignment Scores:				
Pred. No.:				
Score:				
Percent Similarity:				
Best Local Similarity:				
Query Match:				
DB:				
US-10-018-706-2 (1-322) x BH705150 (1-285)				
Oy		136	ThrProSerProVaAlaValaGlnSerSerArgProProVaAlaGlnHisProAlaVal	155
Dd		163	ACACCACATCCACCGCTGTGAACGCCACCAACACACACCAACCAATCCACCACTGTT	104
Oy		156	GlnIlySProThr-----ProProVaIvalValValIlySLySProThProThProPro	173
Dd		103	ACGCCCAACCAACCGCCCGCCAGCTGTGTGAACGCCACCAACCAACCAACCAACCGG	44

Oy	174	ValvalaJngIngnlProAlaproValalaProProValThr	186
Db	43	GTCGTACGCCAGCAGCACCAACCCTCTGTCGTAAACA	5
RESULT 17	AZ187441	735 bp	DNA linear GSS 30-AUG-2000
LOCUS	AZ187441/c		
DEFINITION	SP_1009_A2_D09_SP6E Strongylocentrotus purpuratus, purple sea urchin, sperm genomic BAC library Strongylocentrotus purpuratus genomic clone Plate=1009 Col=18 Row=G, DNA sequence.		
ACCESSION	AZ187441		
VERSION	AZ187441.1	GI:8370620	
KEYWORDS	GSS.		
SOURCE	Strongylocentrotus purpuratus.		
ORGANISM	Strongylocentrotus purpuratus Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa; Echinoidae; Euechinoidea; Echinacea; Echinida; Strongylocentrotidae; Strongylocentrotus. 1 (bases 1 to 735)		
REFERENCE	Cameron,R.A., Mahairas,G., Rast,J.P., Martinez,P., Biondi,T.R., Swartzell,S., Wallace,J.C., Pouscka,A.U., Livingston,B.T., Wray,G.A., Ettensohn,C.A., Lehrach,H., Britten,R.J., Davidson,E.H. and Hood,L. A sea urchin genome project: Sequence scan, virtual map, and additional resources Proc Natl. Acad. Sci. U.S.A. 97 (17), 9514-9518 (2000) Contact: Cameron, RA, Davidson, EH, Hood, L Division of Biology 156-29 California Institute of Technology Pasadena California 91125, USA Tel: (626) 395-8421 Fax: (626) 793-3047 Email: acameron@caltech.edu Plate: 1009 ROW: G column: 18 Seq primer: SP6 Class: BAC ends High quality sequence stop: 735.		
FEATURES	Location/Qualifiers		
source	1..735 /organism="Strongylocentrotus purpuratus" /db_xref="taxon:7668" /clone_xref="plate:1009 Col=18 Row=G" /clone_lib="Strongylocentrotus purpuratus, purple sea urchin, sperm genomic BAC library" /note="Organ: sperm; Vector: Bacc3.6; BAC clones in E-Coli DH10B"		
BASE COUNT	178 a 71 c 311 g 175 t		
ORIGIN			
Alignment Scores:			
Pred. NO.:	0.299	Length:	735
Score:	124.00	Matches:	41
Percent Similarity:	45.99%	Conservative:	22
Best local Similarity:	29.93%	Mismatches:	53
Query Match:	7.43%	Indels:	21
DB:	17	Gaps:	6
US-10-018-706-2 (1-322) x AZ187441 (1-735)			
Oy	80	ThrValserIyslealeaglnArgTyrGlyLeuan-----TrparGluileGlyHis	97
Db	701	ACCAATATCTGCACCTTAACGGTAACCTCTCTCTCTCCACCTCCATACCCCCA	642
Oy	98	IleasnAnleuaenSerSerTyThrIleTyrThrIlgIntPrleutnrlleutripSer	117
Db	641	CTTATATCATAC--CTTCTCTAACCA-----	618
Oy	118	GlyaspneuYsValaAgglunArgserIleserSerGlyValasn-ThralahistHrr	137
Db	617	---CCCTCGCTGTGCACCTTAATACACTAGCCACCACTCTCACCCCCCACACCC	561

```
QY   0serProValAlaVal-----GlnSerArgProValGlnGlnHisproAlaVa 155
      |||          |||         |||         |||         |||
Db   560 CcCTCCcCAGccCcCCTAAcTCAcTGAcGACcCTTCACcATCAAGCCAcCACccCGGT 501

QY   155 lGlnLysProThHProProValValValLyLysProThHProThHProProVal-- 174
      |||          |||         |||         |||         |||
Db   500 cCcCTTCCcCATmCcCccCACccCTcCTcccccccCcCACCACCAcCCAcCATcCC 441

QY   175 -ValGlnGlnProAlaproValAlaProProValThGlnAlAPropheAlathrsLySe 194
      |||          |||         |||         |||         |||
Db   440 cCTcTCcATTAcTAcTCCcCCAAGATCCCCcCTcGCcGGcAGcCCcCTCCcCCcAcCAcCG 381

QY   194 rSerGlyValMetGlnPheArgTyPrProValGlyAlaThrAspProval 210
      |||          |||         |||         |||         |||
Db   380 CTGC-----CTGCATTCTCCcACAACcCTTTCACATGCACGCCCTT 338
```

LOCUS BM913728 1252 bp mRNA linear EST 12-MAR-2000
DEFINITION ACENOCURF 6612844 NIH_MGC_98 Homo sapiens cDNA IMAGE:5477423
5', mRNA sequence.
BM913728
ACCESSION BM913728.1 GI:19364107
VERSION EST.
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 1252)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished. (1999)
COMMENT Contact: Robert Strausberg, Ph.D.

CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
Plate: L1MG195 row: 9 column: 24
High quality sequence step: 398
Location/Qualifiers
1..1252

```

/rtissue type="astrocytoma grade IV, cell line"
/lab host="DH10B (phage-resistant)"
/note="Organ: brain; Vector: pOTB1; Site: 1: XhoI; Site: 2:
EcoRI; cDNA made by oligo-dt priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCAAGAG (5). Library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using Zap-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH-MGC Library. 11 RT (Life Technologies).
BASE COUNT      218 a      568 c      230 g      235 t      1 others
ORIGIN

```

Pred. No.:	0.64	length:	1252
Score:	124.00	Matches:	56
Percent Similarity:	32.77%	Conservative:	22
Best Local Similarity:	23.53%	Mismatches:	84
Query Match:	7.43%	Indels:	76
DB:	14	Gaps:	9

US-10-018-706-2 (1-322) x EM913728 (1-1252)

Qy 21 IiepegivallietnrhCysllleuallaglyCysAlaserlyProThrTyrasn 40

Db	169	ATAATTTGGCATTTTCCAGTTCTGTACGCATC-----TGGGGGTTGGACGCCACGGCCGCTT	222
Oy	41	SerThnSerGlySerGlySerHisArgThnSerGlySerGlyLeuAlaIleGlySer	60
Db	223	ACTCTACGGGAGATC--AGCTGCTCACTGGGGGGACATGGCCCTCTTAGTGT--	273
Oy	61	GlnValIleThnAspSerGlnGlyValProAsnArgTyGlnValIysGlnGlyAspThr	80
Db	274	-----AACAT	279
Oy	81	ValSerLyIleAlaGlnArgTyGlyLeuAsnTrpArgGluIleGlyHisIleAsnAsn	100
Db	280	GTCACACAGTGATGTTTAAATAAATGGCCCGTTAGGGGCTC-----TATTGCAATATAT	330
Oy	101	-LeuAsnSerSerTyThrIleTyThrGlnGlnIntPleuThnLeuThr-----	116
Db	331	GTTTAGGAGCATTTATATATGATTTCCACGACAGGTAC-----ACATATGGCCGTAGAGTAA	384
Oy	117	-SerGlyAspLeu-----	120
Db	385	CACGGGGCCCATTTGATCCACTTTGCACAGACTATTCTGCGCCGCCCTCCGCCCCACCC	444
Oy	121	-----LysValArgGluArgSerIleSerSerGln	130
Db	445	CTCCCCCTCACCGCACGACGCCCCCATCTCCCCCTCGTACGCTCTTCCCCCGCG	504
Oy	130	ValAlaenThrAlaHisThrProSerProValAlaValaGlnSerSer-----	145
Db	505	CACGGCCACACACACACACCCCCCCCCCGCGTACGACGACGACGCGCCCGCCAGTGGCGCC	564
Oy	146	-----ArgProProValGlnGlnHisProAlaValaGlnLysProThrProProVa	162
Db	565	CCACACTCTTACCCCCCTGCCACACCCGCCACCCGCCCATCACCCCCCCACACCCCCC	624
Oy	162	IValaValaValLysLysProThrProThrProProProValaValaGlnGlnProAlaProValaI	182
Db	625	CTCGGACACTCGGGCGCCCGCCCGCCCGCGCCCTTTTACACCGCCCGCCCGCGCTGCG	684
Oy	182	AspProProValThr-GluAlaProPheAlaThrGlySerSerGlyValaIleGlnPheArgT	202
Db	685	CCCCGCCCCCACTTCCCGCCCGCCCGCGCGCCGCCACCGGCGCTTCACGCGCCCCAMA	744
Oy	202	yrProVal-----GlyAlaThrAspProVal	210
Db	745	CTCCCGCTCATCCCAACGACGCGCTCAGAGGCGCAACCGGCGCCGCTG	794

RESULT 19	
BE240905	538 bp mRNA linear EST 16-OCT-2000
DEFINITION	SS0423 <i>Suaeda salsa</i> ZAP cDNA library <i>Suaeda maritima</i> subsp. <i>salsa</i>
ACCESSION	BE240905
VERSION	BE240905
KEYWORDS	cDNA, mRNA sequence.
SOURCE	BE240905.1 GI:9057090
ORGANISM	EST. <i>Suaeda maritima</i> subsp. <i>salsa</i> . <i>Suaeda maritima</i> subsp. <i>salsa</i> . Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Caryophyllales; Caryophyllales; Chenopodiaceae; <i>Suaeda</i> . 1 (bases 1 to 538) Zhang, L., Ma, C. L., Wang, P. D., Sun, Y. F., Zhao, Y. X. and Zhang, H. Expressed sequence tags from a halophyte <i>Suaeda salsa</i> cDNA library Unpublished (2000) Contact: Hui Zhang Key Laboratory of Plant Stress Research The Biology Department of Shandong Normal University No.88, Wenhua East Road, Jinan, Shandong Province, 250014, PRC Tel.: (86)531-2960864 Fax: (86)531-2966954 Email: zhangh@snnu.edu.cn
FEATURES	Location/Qualifiers
Bounce	1..538

BASE COUNT	96 a	227 c	93 g	122 t
ORIGIN				
Alignment Scores:				
Pred. No.:	0.212	Length: 538		
Score:	123.50	Matches: 33		
Percent Similarity:	45.68%	Conservative: 4		
Best Local Similarity:	40.74%	Mismatches: 39		
Query Match:	7.40%	Indels: 5		
DB:	10	Gaps: 3		
US-10-018-706-2 (1-322) x BE240905 (1-538)				
Oy	133	ThralahisthrProserProvalAlaValGlnSerArgProProvalGlnGlnHts	152	
Db	112	TCACACTACATCCCTCCCTCCCTGTTGCCAAGCTCCTCAACACCTCCCTCAAGTCTCC	171	
Oy	153	ProAlaValGlnLysProThProProValValValValLysLysProThProThPro	172	
Db	172	CCACCTGTCCTGCTACCCACCTCCTCAAGTGTCTGCTACCCACCACTGTTGCCACCCCA	231	
Oy	173	ProValValGlnGlnProAlaProValAla-----ProProVal---ThrglnuLabro	189	
Db	232	CCACCTGTTGCTACCCACCTCCTCAAGTGTCTACCCACCTCCAGTGTACCCACCA	291	
Oy	190	PheAlaThrGlySerSerGlyValMetGlnPheAlaGlyrProValGlyAlaThrasnPro	209	
Db	292	GTTGCCACCCACCACTGCTGTCCACCCAGTGTGTACCCCT-----GCTGTGTAGCCCT	345	
Oy	210	Val 210		
Db	346	GTT 348		
RESULT 20				
CNS01SZG	1002 bp DNA linear GSS 12-MAY-2000			
LOCUS	Tetraodon nigroviridis genome survey sequence PUC-ori end of clone			
DEFINITION	170H14 of library G from Tetraodon nigroviridis, genomic survey sequence.			
ACCESSION	AL165877.1 GI:7803615			
VERSION	GSS: genome survey sequence.			
KEYWORDS	Tetraodon nigroviridis.			
SOURCE	Tetraodon nigroviridis.			
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Acanthopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorphae; Tetraodontiformes; Tetraodontidae; Tetraodon.			
REFERENCE	1 (bases 1 to 1002)			
AUTHORS	Roest-Crollius,H., Jallou,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Winkler,P., Brotlier,P., Quettier,F., Saurin,W. and Weissenbach,J.			
TITLE	Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence			
JOURNAL	Unpublished			
REFERENCES	2 (bases 1 to 1002)			
AUTHORS	Roest-Crollius,H., Jallou,O., Dasilva,C., Fizames,C., Fisher,C., Bouneau,L., Billault,A., Quettier,F., Saurin,W., Bernot,A. and Weissenbach,J.			
TITLE	Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis			
JOURNAL	Unpublished			

REFERENCE	3 (bases 1 to 1002)
AUTHORS	Genoscope.
JOURNAL	Direct Submission
COMMENT	Submitted (12-Apr-2000) This sequence is a single read and was generated as part of a large scale clone-and-sequencing project of the Tetradon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetradon .
FEATURES	
source	1. 1002 /organism="Tetradon nigroviridis" /db_xref="taxon:99883" /clone="170H14" /clone_lib="g" /note="Genoscope sequence ID : COAG170DD07SP1-end : PUC-Or1"
BASE COUNT	238 a 315 c 242 g 195 t 12 others
ORIGIN	
Alignment Scores:	
Pred. No.:	0.519 Length: 1002
Score:	123.50 Matches: 66
Percent Similarity:	38.24% Conservative: 25
Best Local Similarity:	27.73% Mismatches: 98
Query Match:	7,40# Indels: 49
DB:	17 Gaps: 14
US-10-018-706-2 (1-322) x CNS01SZG (1-1002)	
OY	104 SerYrThrlleYrThrThcYlgnTrPhleuThrlleTrPserClyAspIeuIysValarg 123
Db	18 AATTACACCATATACAGAGATCAT-----ACGGTGGGATTTGGACATCTTAGAGAG 71
OY	124 -----GluATGserIleSerSerGlyValaIenThralaIheIthr---ProSerProval 140
Db	72 AACGAGGAGAGAGAGACTTCAAAIGTGGCCAGCAAAAGCGTCGCAACTGACCTTAACCAAG 131
OY	141 -----AlaValGlnSerSerArgProProvalGlnGlnIhisProAlaI 155
Db	132 CAGAGGACAGGGGAGGAGCGGTGCTCGGCCACAAAGCGCTTGTAGACAGACAA 191
OY	156 GlnIysProThrProProvalValValIysIysProThrProThrProProvalVal 175
Db	192 CAAGGGCCT-----GACGACCTAAATCTAACCCCGCCAGCA--- 227
OY	176 GlnGlnProAlaProValAlaProProvalThrGlnAla---ProPheAla-----Thr 192
Db	228 ---NACCCAAACCTTCATCGGCGCCTTGGCCAACTGCCGCGAGCTTGTACGCAACAG 284
OY	193 GlySerSerGlyValaIeGlnPheArgTrProValGlyAlaIThrAsnProvalValarg 212
Db	285 TCATCTTCGGGAGATGATCGTTCCAAAGATGMAACGGATCCTGATCCCGCCCGG 344
OY	213 ArgPheGly---ThrAlaThrValAlaGlySerThrValThrSerAsnGlyMetTrpPhe 231
Db	345 AAATWAGGCCACAGCGCTCCATTCGCCGAGC-----GGTGGGGGCAATTTGGAGC 395
OY	232 SerGlyArgAspGlyAspIeuIleAsnAla-----SerAsnAlaGlyThrValIle 248
Db	396 ATTGGGAAGAGTGCACACTTGTGTCTACTACGTTCAACAGCAGCAGAGGCCAGACAGT 455
OY	249 GlnAlaIlePheIAsn-----MetAspGlyAlaSerIleVal 260
Db	456 TTAGTGTATCAAGACTGTTTGGATTTCTGATGTGATGATAGCGGCCAGATCACA 515
OY	261 IleGlnIhis---ThrAsnGlyPheValSerSerTrpIleHisIleIysAspAlaGlnVal 279
Db	516 GCTCAATATGGCGTGAATGGTTTTTCGTGTGATCCCTGACCTAAACGCTTCGGCGATC 575
OY	280 LysThrGlyAspThrValargThrGlyGlnArgIleAlaSerMetIysAsnGlnProSer 299
Db	576 TCC-----GCCAGTAACTGGATCAAGCTT 602

```
Oy      300 GlyAlaIalLeuPheGluPheArgIleSerArgAsnGlyValTyrValaspPro 317
          |||||
Db      603 TCGCAACAGCTGTGGATTGAGAGAGACTTGACGCGTTGAGCGGCATCCC 656
```

RESULT 21	Bj220947/c	669 bp	mRNA	linear	EST 04-APR-2002
LOCUS	Bj220947	669	bp		
DEFINITION	cdna clone w122109 3', mRNA sequence				
ACCESSION	Bj220947				
VERSION	Bj220947.1				
KEYWORDS	EST				
SOURCE	Bread wheat.				

REFERENCE
1 (bases 1 to 689)
Ogihara, Y. and Murai, K.
Expressed genes in *Triticum aestivum*
Unpublished (2002)
JOURNAL
COMMENT
Contact: Tadasu Shi-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel.: 81-559-81-6855
Fax: 81-559-81-6855
Email: tshin@genes.nig.ac.jp.
FEATURES
location/Qualifiers
1..689

BASE COUNT	160 a	157 c	165 g	187 t
ORIGIN				
Alignment Scores:				
Pred. No.:	0.336		Length:	689
Score:	123.00		Matches:	58
Percent Similarity:	38.91%		Conservative:	27
Best Local Similarity:	26.61%		Mismatches:	86
Query Match:	7.37%		Indels:	47
DB:	13		Gaps:	10

US-10-018-706-2 (1-322) x BJ220947 (1-689)

QY	68	GLYValProAsnArgTyrGlnValLysGlnGlyAspThrValSerLysLeuIaGlnArg	87
Db	689	GGGATCCCCAAT-----GTGCAGCCCATACATCAACGMAAGAGCTCTCTCAACCG	639
QY	88	TyrgLYLeuAsnTrpArgGlnIleGlyHisIleAsnAsnLeuAsnSerSerTyrrHille	107
Db	638	ATGGGGCTC-----ACCAAG	624
QY	108	TyrrThGlyGlnTrpLeuThrLeuTrpSerGlyAspLeuLysValArgGlnArgSerIle	127

```

Db      622 TTCTCCAGGCGCTTCAAAAAGAGTGG-----5373
Qy      128 SerSer-GlyValaantPrAlaIethProSerProValaIaValGlnSerSerArgPr 147
      128 |||||::: |||||:::
Db      596 CCAGGACACAGGTTGATGCTCCACAGACAGACAAACTACAAAAGTCAAAACAACACACC 5373
Qy      147 oProValGlnGlnHisPro--AlaValGlnLysProThrProProValValValLys 166
      147 ::::: |||||:::
Db      536 TTTCGATTGGATTACCTCGAGGACCTTCGGTTCCACACCACTGCAT-----4853
Qy      166 sLysProThrProThrProProValValGlnGln--ProAlaProValAlaProProVa 185
      166 |||||::: |||||:::
Db      484 ----CCGGTTCCAAACCAACTGTACTTCAACTTTCCAAACCAACTGGGCCACCACT 4293
Qy      185 lThrguAlaProPhaIaethGlySerSerGlyValMetGlnPheArgTyr---ProVa 204
      185 |||||::: |||||:::
Db      428 TCGTGAAGAACCTTCTACGATGGCCATATGAGATCCGTCCTADAGATTCAACACAGT 3683
Qy      204 lGlyAlaThrAsnProValValArgArgPheGlyThrAlaThrValAlaGlySerThrVa 224
      204 |||||::: |||||:::
Db      368 TGAATGTAACCAATCCACCCACGAGAGTGGTGTGATGGCGCTAAACTTGCMAAACCTGT 3073
Qy      224 lThrSerAsnGlyMetTyrPheSerGlyArgAspGlyAsnLeuIleAsnAlaSerAsnAl 244
      224 ::::: |||||:::
Db      308 AGAA-----AAATGAGCCCATCAACGACAGAGTGTATACACCCCTC 2673
Qy      244 aGlyThr--ValIleGlnAlaAspHisAsnMetAspGlyAlaSerIleValIleGlnH 263
      244 |||||::: |||||:::
Db      266 TGGTACTGTGTGATTTGCTGGATGTCCTGTGGAGAAGAGCTGCATTCATGTGTGCA 2073
Qy      263 eThraAsnGlyPheValSerSerTyrIleHisIleLysAsnAlaGlnValLys 280
      263 ::::: |||||:::
Db      206 CATGGCTGGTTGCATGCTCTGC-----TTGAAGATATCAAGTGCAGG 164

```

RESULT 22
LOCUS AU268480
DEFINITION AU268480 VS Dictyostelium discoideum cDNA clone VS1307 5', mRNA sequence.
ACCESSION AU268480
VERSION
KEYWORDS
SOURCE
ORANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

500 bp mRNA linear EST 10-MAY-2002

AU268480 VS Dictyostelium discoideum cDNA clone VS1307 5', mRNA sequence.
AU268480
AU268480.1 GI:20527278
EST.
Dictyostelium discoideum.
Dictyostelium discoideum.
Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.
1 (Bases 1 to 500)
Tsunehira,H., Morio,T., Saito,T., Koriiki,E., Ochiai,H., Maeda,M., Takeuchi,I., Kohara,Y. and Tanaka,Y.
Population analysis of cDNAs from unicellular and multicellular stages of Dictyostelium discoideum
Unpublished (2002)
Contact: Hideko Tsunehira
Institute of Biological Sciences
University of Tsukuba
1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
Tel: 81-298-53-4664
Fax: 81-298-53-6614
Email: hideko@biol.tsukuba.ac.jp.

source	1.500	/organism="Dictyostelium discoideum"	
		/strain="AX4"	
		/db_xref="taxon:44689"	
		/clone="VSI307"	
		/clone_11b="VS"	
		/sex="mat A"	
		/dev_stage="vegetative"	
BASE COUNT	212 a	151 c	50 g
ORIGIN			85 t
			2 others
Alignment Scores:			
Pred. No.:	0.235	Length:	500
Score:	122.50	Matches:	41


```

/clone="PTB-040D10.F"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT      266 a      412 c      442 g      219 t      24 others
ORIGIN

```

Alignment Scores:

```

Pred. No.:      0.995      Length:      1363
Score:          122.50      Matches:      69
Percent Similarity: 34.38%      Conservative: 41
Best local Similarity: 21.56%      Mismatches: 141
Query Match:      7.34%      Indels:      69
DB:              17      Gaps:      15

```

US-10-018-706-2 (1-322) x AG054509 (1-1363)

```

Qy 36 LysProThrTyraSerThrSerGlySerGlySerHisArgThrSerGlySerGly 55
Db 248 AAACCGACGAGAAAAAGAAAGAAACATMAAAAGAGAGACGGAGGGGAGCGGC 307
Qy 56 LeuAlaIleGlySerGlnValIleThrAspSerGlnIValPro---AsnArgTyrGln 74
Db 308 CGC-----GGCGGGCGAGGGGTGCGGGAACCGCGCGCATCTCGCGGCGCATCGCG 361
Qy 75 ValIleGlnIleAspThrValSerIleAlaGlnArgTyrGlyLeuAsnTrpArgIu 94
Db 362 GAGATGACGGCGCGCGCGCGCATCTCGCGCATGATCGCGCGCGCATGTGCGCGCG 421
Qy 95 IleGlyHisIleAsnLeuAsnSerSerTyrThrIleTyrThrGly-----GlnTrp 112
Db 422 CGCGGCTTATATAGGATACGCCCTTCTCATCATGTTATGCTGAGTATGTGCTG 481
Qy 113 LeuThrLeuTrpSerGlyAspLeuIysValArgIleAspSerIleSerGly----- 130
Db 482 TCACCTCCCTTGGCTACCTATTCGTCGTCACGTCGCTCACCATMAACGGAGCGGC 541
Qy 131 -----ValAsnThrAlaHisThrProSerProValAlaValGlnSerSerArg 146
Db 542 GATGAATACGTATGTGTGCTGGCGGTGCGGGCGCGCGCGCGCGCGCGCAT 601
Qy 147 ProProValGlnGlnHisProAlaValGlnIleProThrProProValIValIValIys 166
Db 602 GCTCG-----CCCGAGTTCGCCGTGCGGTCCCGCATCTCTGTTTCG 649
Qy 167 LysPro---ThrProThrProProValIValGlnIleProAlaProValAlaProProVal 185
Db 650 TTCCCTTATTAACCGCGCGCGGTACTATGACCATCCACACCCCAACCGCGGAACATC 709
Qy 186 ThrGlnAlaProPheAlaThrGlySerSerGlyValMetGlnPheArgTyrProValGly 205
Db 710 AATGACAGTGCAGACCGCGGTGTTCCGA---CGCGGTGCGGACCTCCCTGCGAC 766
Qy 206 AlaThrAsn-----ProValIleArgArgPheGlyThrAlaThrValAlaGlySerThr 223
Db 767 GCCTCGGTCATCCACGCTTATGTCGCATATACGTCGCGGAGGAGCGCT 826
Qy 224 ValThrSerArgnGlyMetTrpPheSerGlyArgAspGlyValPheIleAsnIleAsn 243
Db 827 -----TCCCTCTGCTCATCCGAC 847
Qy 244 AlaGlyThrValIleGlnAla-----AspHisAsnMetAspGly 256
Db 848 TGCCTGCGGTGCGGCTTCTTATGCGAGAGAGCGCGCGCGCGCGCGCGCG 907
Qy 257 AlaSerIleValIleGln----- 262
Db 908 GCGTCGCTGCTTGGCTGTCTTACCGCTCAAACTCGGGCGCGATGTTTCCACTGT 967
Qy 263 -----HisThrArgnIle-----PheValSerSerTyrIle----- 272
Db 968 TGTACATCTGACGATCGACGAGCGCGGTGCTCCGCGCATCACGATCCGACGAGTAA 1027

```

```

Qy 273 -----HisIleLeuAspAlaGlnValIleThrGlyAspThrValArgThrGlyGlnArg 290
Db 1028 GATCGGCGCTTACCGACCGACCGACGCGTGG-----ACAGCANATATCTCATCGCT 1078
Qy 291 IleAlaSerMetLysArgnGlnProSerGlyValAlaIlePheGluPheArgIleSerArg 310
Db 1079 TGTGAGCGCGATATATAT---CCGACGGCGCGTGGCTGTGTGTCGGTACGATAGA 1135

```

RESULT 25

```

BM109778      450 bp      mRNA      linear      EST 26-NOV-2001
LOCUS      EST557314 potato roots Solanum tuberosum cDNA clone cPRO5C4 5' end,
DEFINITION      mRNA sequence.
VERSION      BM109778
KEYWORDS      BM109778.1 GI:117070945
SOURCE      EST.
ORGANISM      potato.

```

Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asterales; euasterids I; Solanales; Solanaceae; Solanum.
1 (bases 1 to 450)
Van der Hoeven, R., Sun, H., Karameyeva, S.A., Tsai, J., Van Aken, S.,
Utterback, T., Chiemingo, A., Bougri, O., Buell, C.R., Romning, C.,
Tankley, S. and Baker, B.
Generation of ESTs from potato roots
Unpublished (2001)
Contact: Research Genetics, Libraries Division
Tel.: 1-800-711-6195

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Email: cdna@resgen.com
For clone info, please contact Research Genetics, Libraries
Division tel 1-800-711-6195, email cdna@resgen.com
Seq primer: T3

FEATURES

```

source
location/Qualifiers
1..450
/organism="Solanum tuberosum"
/cultivar="Kennebec"
/db_xref="taxon:4113"
/clone="cPRO5C4"
/clone_lib="potato roots"
/tissue_type="roots"
/dev_stage="in vitro grown stem cuttings"
/lab_host="SOLR"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; supplier: Cornell University, Tankley lab;
sequencing: The Institute for Genomic Research. Roots were
isolated from in vitro grown stem cuttings on CW medium.
Roots were isolated two weeks after placing the stem
cuttings from in vitro grown plants on medium."
BASE COUNT      125 a      190 c      30 g      105 t
ORIGIN

```

Alignment Scores:

```

Pred. No.:      0.225      Length:      450
Score:          122.00      Matches:      31
Percent Similarity: 48.19%      Conservative: 9
Best local Similarity: 37.35%      Mismatches: 31
Query Match:      7.31%      Indels:      12
DB:              13      Gaps:      5

```

US-10-018-706-2 (1-322) x BM109778 (1-450)

```

Qy 136 ThrProSerProValAlaValGlnSerSerArgProProValGlnGlnHisProAlaVal 155
Db 160 ACTTCAACCAACCACTTATGTC-----CATCCACCAATCATCTCAAAACCA----- 204
Qy 156 GlnIleProThrProProValIVal-----ValAlaIleLysIlePro---ThrProThr 171
Db 205 CCATCAACCAACCACTTATGTTTACCCCTTATGTTTATCCACCAATCAACACCA 264
Qy 172 ProProValIValGlnIleProAlaProValAlaIleProProValIleThrGlnAlaProPheAla 191
Db 265 CCACCTATGTGTACCTCATCATTCATCAACACCACTTATGTTTCTCCACCTTTGTC 324

```


Qy 192 ThrGlySerSerGlyValMetGlnPheArgTyr-----ProValGlyAlaThrAsn 208
 Db 325 CCACATCCCTCCGGGTGTATACACACACCTTACGCAAGTCTCCGGGTGTACTCCA 384
 Qy 209 ProValVal 211
 Db 385 CCCATTAATT 393
 RESULT 26
 AG159588/c 1140 bp DNA linear GSS 09-JAN-2002
 LOCUS DEFINITION Pan troglodytes DNA, clone: RP43-025C14.TU, genomic survey
 sequence.
 ACCESSION AG159588
 VERSION AG159588.1 GI:16689266
 KEYWORDS GSS.
 SOURCE Pan troglodytes male lymphocytes DNA, clone: lib:RP43-43 Chimpanzee
 Male BAC Library clone:RP43-025C14.TU.
 ORGANISM Pan troglodytes
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.
 REFERENCE 1 Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
 Auteurs Tokoki, Y., Matanabe, H. and Sakaki, Y.
 BAC end sequences of Library RP43-43
 JOURNAL Unpublished
 TITLE 2 (bases 1 to 1140)
 AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
 Tokoki, Y., Matanabe, H. and Sakaki, Y.
 JOURNAL Direct Submission
 Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
 1-7-22 Suehiro-chou, Tsunumi-ku, Yokohama, Kanagawa 230-0045, Japan
 (E-mail:chimbee@sc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
 Tel:81-45-503-9111, Fax:81-45-503-9170)
 Clones are derived from the chimpanzee BAC library RP43-43 This BAC
 end was generated during the R&D process and may have higher chance
 of clone tracking errors.
 COMMENT PRIMERS
 Sequencing: TU
 LIBRARY
 Vector : pBACe3.6
 R.Site 1 : EcoRI
 R.Site 2 : EcoRI.
 FEATURES
 source
 1. 1140
 Location/Qualifiers
 /organism="Pan troglodytes"
 /db_xref="taxon:9598"
 /clone="RP43-025C14.TU"
 /sex="male"
 /cell_type="lymphocytes"
 /clone_lib="RP43-43 Chimpanzee Male BAC Library"
 BASE COUNT 34 a 347 c 476 g 282 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 0.949 Length: 1140
 Score: 121.50 Matches: 47
 Percent Similarity: 41.45% Conservative: 16
 Best Local Similarity: 30.92% Mismatches: 58
 Query Match: 7.28% Indels: 31
 DB: 17 Gaps: 7
 US-10-018-706-2 (1-322) x AG159588 (1-1140)
 Qy 137 ProSerProValAlaValAlaGlnSerSerArgProProValGlnGlnHisProAlaValGln 156
 Db 1121 CCGGGGGCCACCC-----CCCGGGGGAACCCCGACCCCGCCCGCCCGC 1080
 Qy 157 LysProThrProProValAlaValAlaValLysLysProThrProThr-----ProPro 173
 Db 1079 CGACCAAGCCCGCCACCTCCCGCTCTCCCGCGGAACCAAGTCCCGGCGCGCCCGCC 1020

Qy 174 ValValGlnGln-ProAla---ProValAlaProProValThrGluAlaProPheAlaTh 192
 Db 1019 CAGATAGGGGACACCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 960
 Qy 192 rGlySer-----SerGlyValMet-GlnPheArgTyrProValGlyA 206
 Db 959 GGGCAAAACATCAGCCACCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCG 900
 Qy 206 lathraenProValAlaArgArgPheGlyThrAlaThrValAlaGlySerThr---ValT 225
 Db 899 CGCGCACCCCAACCCCGCAAGAACAGGCGACGTCCACCCCGCCCGCCCGCCAGATGA 840
 Qy 225 hrSerAsn-----GlyMetTrpPheSerGlyArgAspGlyAspL 238
 Db 839 CCGCAACACAGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 780
 Qy 238 euileAmAlaSerAsnAlaGlyThrValIleGlnAlaAspHisAspMetAspGlyAla- 257
 Db 779 CGCGGAACNMAACATCGCGCGGGCGAGAACACAGAGCGGCAACACACCGCGCGCCAT 720
 Qy 258 -----SerIleValIleGlnHis 263
 Db 719 GAAACAAACAACACACATCGCGCGCGCGCCACAT 688
 RESULT 27
 B1174098 508 bp mRNA linear EST 09-JUL-2001
 LOCUS OSTF009E10_1 AD-wrmcDNA Caenorhabditis elegans cDNA similar to
 DEFINITION F13D12.3, mRNA sequence.
 ACCESSION B1174098
 VERSION B1174098.1 GI:14639886
 KEYWORDS EST.
 SOURCE Caenorhabditis elegans.
 ORGANISM Caenorhabditis elegans.
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea
 ; Rhabditidae; Pelodidae; Caenorhabditis.
 REFERENCE 1 (bases 1 to 508)
 AUTHORS Reboul, J., Vaglio, P., Tzellas, N., Thierry-Mieg, N., Moore, T.,
 Jackson, C., Shin, I.T., Kohara, Y., Thierry-Mieg, D., Thierry-Mieg, J.,
 Lee, H., Hiti, U., Doucette-Stamm, L., Hartley, J.L., Temple, G.F.,
 Brasch, M.A., Vandenhaute, J., Lamesch, P.E., Hill, D.E. and Vidal, M.
 Open-reading-frame sequence tags (OSTs) support the existence of at
 least 17,300 genes in C. elegans
 Nat. Genet. 27 (3), 332-336 (2001)
 JOURNAL 21135099
 MEDLINE
 COMMENT Contact: Reboul J, Vaglio P
 Marc Vidal Laboratory
 Dana Farber Cancer Institute
 44 Binney Street, Boston, MA 02115, USA
 Tel: 617 632 5180
 Fax: 617 632 2425
 Email: Jerome.Reboul@dfci.harvard.edu
 Sequence tag of Gateway entry clones. The primers used were
 designed on the predicted protein encoding ORF. C. elegans ORFome
 cloning project : Contact jerome_reboul@dfci.harvard.edu or
 philippe_vaglio@dfci.harvard.edu
 POLY(A)=No.
 FEATURES
 source
 1. 508
 Location/Qualifiers
 /organism="Caenorhabditis elegans"
 /strain="N2"
 /db_xref="taxon:6239"
 /clone_lib="AD-wrmcDNA"
 /sex="Hermaphrodite and male"
 /tissue_type="whole animal"
 /dev stage="mixed stage"
 /note="The AD-wrmcDNA library was generated with poly(A)+
 RNA isolated from both hermaphrodite and male N2 worms of
 all larval stages, embryos, adults and dauers and the
 subsequent generation of cDNAs by poly(A) priming. The
 cDNAs were cloned into pPC86"
 BASE COUNT 141 a 148 c 104 g 115 t

ORIGIN

Alignment Scores:

Pred. No.: 0.366 Length: 508
 Score: 120.50 Matches: 54
 Percent Similarity: 34.54% Conservative: 13
 Best Local Similarity: 27.84% Mismatches: 73
 Query Match: 7.22% Indels: 55
 DB: 13 Gaps: 9

US-10-018-706-2 (1-322) x B1174098 (1-508)

```

Qy 28 CyeleleuallaglyCysala-----SerlyspProthryTyranser 41
Db 3 TGCCTCATTTTCATTCACACGCGATTTCGTAAGACTCGAGAGAGCCTACATGTC 62
Qy 42 ThserglyserglySerhiSarghmSerGlySerGlyLeuallleglyserGln 61
Db 63 AGCAGT-----TCATATCTAGGCCCAAGTTACGAGGA-----95
Qy 62 ValIethrAspSerGlnGlyValProAsnAlyGlyGlnValIysGlnIysAspThrVal 81
Db 96 -----GACCAATGGGCTCAGAAATTAATTAATTAATTCATATCAAAACAAATGG-- 143
Qy 82 SerlySleAlaGlnAlyGlyLeuAsnTPArgGluIleGlyHisIleAsnAsnLeu 101
Db 144 -----AATCAACAAACCAAGAAATTCGATCAATATGGA-----AACAGA 182
Qy 102 AsnSerSerTyThrIleTyThrGlyGlnTPLeuThryLeuThrySerGlyAspLeuIys 121
Db 183 GGTAGTCATATGATGATCTCATACGGCGGCC--TTCAGATTAATGGCCACACAGCCAAAG 241
Qy 122 ValArgGlnAlySerIle-SerSerGlyValAsnThrAlaIsthrProSerProValAl 141
Db 242 CTACTCCCAACCAAAATCTCTATATGGA-----CGCCCAAGAACCCACCTA 286
Qy 141 aValGlnSerSerAspPro-----ProValG1 150
Db 287 CTCCTCATATTCGGGCTTACGCTTCATATGGAAGACCACTCACTTATCTCACTCA 346
Qy 150 ngInIserProAlaValGlnIysProthProProValValVal-----164
Db 347 GCCAGGTCCAGGAGGTCAGCTGTGATCTCCAGTTCATCACTCAAGCCAAATGTCGCC 406
Qy 165 -----ValIysIysProthProthProthProthProthProthProthProth 180
Db 407 TCCAGTTCCAGAAATTTCTAGCCAGTTCAGAGCCACCA--ATCCAGCTCCAGGCC 463
Qy 180 oValAlaProProValThrGlnAlaProPhealathrGly 193
Db 464 TGTTCCTCCACCAACCATCTGATCTCATATTCGAGGA 503

```

RESULT 28
 CNS03NOR/c 969 bp DNA linear GSS 17-MAY-2000
 LOCUS Tetradon nigroviridis genome survey sequence T7 end of clone
 DEFINITION 040D21 of library G from Tetradon nigroviridis, genomic survey
 sequence.
 ACCESSION AL252396
 VERSION AL252396.1 GI:7973408
 KEYWORDS GSS: genome survey sequence.
 SOURCE Tetradon nigroviridis
 ORGANISM Tetradon nigroviridis
 Bacteria; Mollusca; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 Acanthomorphia; Acanthopterygii; Perciformes; Tetraodontiformes;
 Tetraodontidae; Tetradon.
 1 (bases 1 to 969)
 Roest-Crollius/H., Jallion/O., Dasilva/C., Bouneau/L., Fisher/C.,
 Bernot/A., Fizames/C., Wincker/P., Brothier/P., Quetier/F.,
 Saurin/W. and Weissenbach/J.
 Human gene number estimate provided by genome wide analysis using
 Tetradon nigroviridis DNA sequence

JOURNAL
REFERENCE
AUTHORS

Unpublished
 2 (bases 1 to 969)
 Roest-Crollius/H., Jallion/O., Dasilva/C., Fizames/C., Fisher/C.,
 Bouneau/L., Billault/A., Quetier/F., Saurin/W., Bernot/A. and
 Weissenbach/J.
 Characterization and repeat analysis of the compact genome of the
 freshwater pufferfish Tetradon nigroviridis

JOURNAL
TITLEJOURNAL
AUTHORSJOURNAL
COMMENT

FEATURES

source

1. 969
 /organism="Tetradon nigroviridis"
 /db_xref="taxon:99883"
 /clone="040D21"
 /clone_11b="G"
 /note="Genoscope sequence ID : C0B6040CB11b1-end : T7"
 BASE COUNT 215 a 198 c 398 g 151 t 7 others

ORIGIN

Alignment Scores:
 Pred. No.: 0.926 Length: 969
 Score: 120.50 Matches: 30
 Percent Similarity: 39.29% Conservative: 3
 Best Local Similarity: 35.71% Mismatches: 30
 Query Match: 7.22% Indels: 21
 DB: 17 Gaps: 2

US-10-018-706-2 (1-322) x CNS03NOR (1-969)

```

Qy 135 HisThrProSerProValAlaValGlnSerSerArgProProProValGlnGlnHisProAla 154
Db 606 CACCAAGCTCTCTTCATACAGCAAGCCCGCCCTCTCGAACCACACACACTGCA 547
Qy 155 -----ValGlnIysProthProthProthProthProthProthProthProth 166
Db 546 CACGTGCTGCGCTGTTGTCACCAAGTGACCTGCACCCCGCTGCACCTTACT 487
Qy 167 LysProthProthProthProthProthProthProthProthProthProthProth 186
Db 486 CCACCTGCACCTTACTGCTCACCCTGACCTGCACCTGACCTGACCTGACCTGACCT 427
Qy 187 GluAlaProPhealathrGlySerSerGlyValMetGlnPheArgTyProValGlyAla 206
Db 426 GTAGCTTCA-----CTGTAGCTTCA 406
Qy 207 ThrAsnProVal 210
Db 405 CTGCACCTGTA 394

```

RESULT 29
 AL377972 365 bp mRNA linear EST 03-AUG-2000
 LOCUS MBB35B07F1 MBB3 Medicago truncatula cDNA clone MBB35B07 T3, mRNA
 DEFINITION sequence.
 ACCESSION AL377972
 VERSION AL377972.1 GI:9677724
 KEYWORDS EST.
 SOURCE Medicago
 ORGANISM Medicago
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eustoids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;
 Medicago.
 1 (bases 1 to 365)
 Journef,E.P., Crepeau,H., van-Tuinen,D., Gouzy,J., Jallion,O.,
 Nebel,A., Carreau,V., Chataigner,O., Kahn,D., Glaninazzi-Pearson

TITLE
JOURNAL
COMMENT

'V. and Gamas, P.
Medicago truncatula ESTs from Sinorhizobium meliloti-induced root nodules
Unpublished (2000)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqef@genoscope.cns.fr, Web : www.genoscope.cns.fr
Contact : Pascal Gamas and Etienne-Pascal Journet, Laboratoire de Biologie Moléculaire des Relations Plantes-Microorganismes, CNRS-INRA, BP 27 31326 Castanet-Tolosan Cedex, France (Email : Mt-est@toulouse.inra.fr Website : http://sequence.toulouse.inra.fr/Mttruncatula.html).
Location/Qualifiers

FEATURES
source

1. 365
/organism="Medicago truncatula"
/cultivar="Uemalong"
/db_xref="taxon:3880"
/clone="M1B35B07"
/clone_lib="M1B3"
/tissue_type="symbiotic root nodules"
/dev_stage="harvested 4 days post inoculation with Sinorhizobium meliloti"
/note="Vector: pBluescript PSK, Site 1: EcoRI; Site 2: XhoI; Plants were grown in an aeroponic chamber on nitrogen-rich medium for 21 days. Three days before inoculation with Sinorhizobium meliloti, the medium was replaced by N-free medium. Root nodules (+ short adjacent root segments) were harvested 4 days post inoculation. cDNA was prepared from polyA+ enriched RNA. The cDNA was directionally ligated into Uni-Zap XR vector from Stratagene and packaged using GigaPack Gold packaging extracts. Plasmids containing cDNA inserts were mass-excised from phage stocks using Exsacit helper phage and propagated in SOLR cells. Clone ordering and sequencing was performed by the Centre National de Sequencage (Genoscope, Evry, France)."
BASE COUNT 111 a 129 c 36 g 89 t
ORIGIN

Alignment Scores:
Pred. No.: 0.253 Length: 365
Score: 120.00 Matches: 26
Percent Similarity: 63.79% Conservative: 11
Best Local Similarity: 44.83% Mismatches: 13
Query Match: 7.19% Indels: 8
DB: 9 Gaps: 3

US-10-018-706-2 (1-322) x AL377972 (1-365)

QY 137 ProSerProValAlaValGlnSerSerArgProProValGlnGlnHisProAlaValGln 156
DB 83 CTTTCTCCACCATGTGTAATCA-----CCACCTTATCAATCAACCACTTATGTTAG 136
QY 157 LysProThrProProValAla-----ValValLysLysProThrProThr 171
DB 137 GCACCTTCTCCCTCTAGTGAACCACTCCACTTATGTAATCACT---CCTTCT 193
QY 172 ProProValValGlnGlnProAlaProValAlaProProValThGlnAlaPro 189
DB 194 CCTCTTGTGTAAGAAACACCACTTATCAATCAACCACTTATGTAAGGACCA 247

RESULT 30
LOCUS AMS86923 511 bp mRNA linear EST 07-SEP-2000
DEFINITION ESTJ18546 MHAM Medicago truncatula/Glommus versiforme mixed EST
ACCESSION AMS86923
VERSION AMS86923.1 GI:7266437
KEYWORDS EST
SOURCE Medicago truncatula/Glommus versiforme mixed EST library.
ORGANISM Medicago truncatula/Glommus versiforme mixed EST library
Eukaryota; mixed EST libraries.

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1 (bases 1 to 511)
Harrison, M.J., Liu, J., Peng, H., Gonzales, M., Ellis, L., Town, C.D., Bowman, C.L., Craven, M.B., Hansen, T.S., Holt, I.E. and Fraser, C.M.
ESTs from roots of Medicago truncatula after colonization with Glommus versiforme
Unpublished (2000)
Contact: Maria J. Harrison
Plant Biology Division
The Samuel Roberts Noble Foundation
2510 Sam Noble Parkway, Ardmore, OK 73401, USA
Tel: 580-223-5810
Fax: 580-221-7380
Email: mjharrison@noble.org
Noble EST name: N253964e
TIGR sequence name: MTDAV15TK
More information is available at:
http://chrystle.tamu.edu/Medicago/
Seq primer: SKmod (CTA gaa cta gtc gat cc).
Location/Qualifiers

FEATURES
source

1. 511
/organism="Medicago truncatula/Glommus versiforme mixed EST library"
/cultivar="Medicago truncatula genotype A17"
/db_xref="taxon:119092"
/clone="pMHAM-56C5"
/clone_lib="MHAM"
/tissue_type="roots colonized with Glommus versiforme"
/dev_stage="Roots harvested at 10, 17, 22, 31 and 38 days post-inoculation with Glommus versiforme. The library was made from a mixture of RNA from each of these stages."
/lab_host="E. coli strain XL0LR"
/note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2: XhoI; cDNA was prepared from polyA+ enriched RNA from roots harvested at 10, 17, 22, 31 and 38 days post-inoculation with Glommus versiforme. The cDNA was directionally ligated into the UniZap XR vector from Stratagene and packaged using GigaPack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-Zap phage using Ex-sacit helper phage and propagated in XL0LR cells."
BASE COUNT 158 a 169 c 59 g 125 t
ORIGIN

Alignment Scores:
Pred. No.: 0.41 Length: 511
Score: 120.00 Matches: 26
Percent Similarity: 63.79% Conservative: 11
Best Local Similarity: 44.83% Mismatches: 13
Query Match: 7.19% Indels: 8
DB: 10 Gaps: 3

US-10-018-706-2 (1-322) x AMS86923 (1-511)

QY 137 ProSerProValAlaValGlnSerSerArgProProValGlnGlnHisProAlaValGln 156
DB 208 CTTTCTCCACCATGTGTAATCA-----CCACCTTATCAATCAACCACTTATGTTAG 261
QY 157 LysProThrProProValAla-----ValValLysLysProThrProThr 171
DB 262 GCACCTTCTCCCTCTAGTGAACCACTCCACTTATGTAATCACT---CCTTCT 318
QY 172 ProProValValGlnGlnProAlaProValAlaProProValThGlnAlaPro 189
DB 319 CCTCTTGTGTAAGAAACACCACTTATCAATCAACCACTTATGTAAGGACCA 372

RESULT 31
LOCUS B0147993 650 bp mRNA linear EST 24-APR-2002
DEFINITION NR049809FL1P1070 Developing flower Medicago truncatula cDNA clone
ACCESSION B0147993
VERSION B0147993
KEYWORDS EST
SOURCE B0147993.1 GI:20285052

SOURCE
ORGANISM
Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.

REFERENCE
AUTHORS
1 (bases 1 to 650)
Flores-Jerez, I., Scott, A.D., Harris, A.R., Gonzalez, R.A., Bell, C.J., Flores, H.R., Iman, J.T., Weller, J.W. and May, G.D.
Expressed Sequence Tags from the Samuel Roberts Noble Foundation
Medicago truncatula flower library
Unpublished (2001)

JOURNAL
COMMENT
Contact: May GD
Plant Biology Division
The Samuel Roberts Noble Foundation
2510 Sam Noble Parkway, Ardmore, OK 73402, USA
Tel: 580 221 7391
Fax: 580 221 7380
Email: gdmay@noble.org
Insert length: 650
Plate: 049 Row: 2 Column: 08
Std Error: 0.00
Seq primer: TCACACAGGAAACGCTATGAC.

FEATURES
source
1..650
Location/Qualifiers
/organism="Medicago truncatula"
/db_xref="taxon:3880"
/clone="NF049E09FL"
/clone_1ib="Developing flower"
/tissue_type="Developing flowers"
/dev_stage="Developmentally pooled. Contains a mixture of very young, developing, fully-opened flowers and flowers in early transition into pods."
/note="Vector: Lambda Zap; cDNA was prepared from polyA+ enriched, pooled samples of equivalent amounts of total RNA from very young, developing, fully-opened flowers and flowers transitioning into pods. The cDNA was directionally ligated into the Uni-Zap XR vector (Stratagene) and packaged using the Gigapack III Gold packaging extracts. Phagemids containing cDNA inserts were in vivo excised from the recombinant Uni-Zap XR vector using Exassist helper phage and the E. coli strain XLI-Blue MRF (Stratagene). Excised plasmids were plated using SQR cells."

BASE COUNT
ORIGIN
197 a 198 c 77 g 175 t 3 others

Alignment Scores:
Pred. No.: 0.579 Length: 650
Score: 120.00 Matches: 26
Percent Similarity: 63.79% Conservative: 11
Best Local Similarity: 44.83% Mismatches: 13
Query Match: 7.19% Indels: 8
Gaps: 3

US-10-018-706-2 (1-322) x BQ147993 (1-650)

QY 137 ProSeProVaAlaValaGlnSerSeraGProProVaAlaGlnHisProAlaValaGln 156
Db 311 CCTCTCCACCACTGTAATAATCA-----CCACCTTATCAATCAACCACTTATGTAAG 364
QY 157 LysProThProProVaAlaVala-----ValaValaLysProThProThr 171
Db 365 GCACCTTCTCCCTCTAGTGAACAACCACTCCACCTATTTGTGAATCACT--CCTTCT 421
QY 172 ProProVaValaGlnGlnProAlaProVaAlaProProVaAlaThGlnAlaPro 189
Db 422 CCTCTTGTGTGAACAACCACTTATCAATCAACCACTTATGTAAGGACACCA 475

RESULT 32
BI271940 710 bp mRNA linear EST 18-JUL-2001
LOCUS NF016F08FL1074 Developing flower Medicago truncatula cDNA clone
DEFINITION NF016F08FL 5', mRNA sequence.

ACCESSION
VERSION
BI271940
KEYWORDS
EST.
SOURCE
ORGANISM
Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.

REFERENCE
AUTHORS
1 (bases 1 to 710)
Flores-Jerez, I., Scott, A.D., Harris, A.R., Gonzalez, R.A., Bell, C.J., Flores, H.R., Iman, J.T., Weller, J.W. and May, G.D.
Expressed Sequence Tags from the Samuel Roberts Noble Foundation
Medicago truncatula flower library
Unpublished (2001)

JOURNAL
COMMENT
Contact: May GD
Plant Biology Division
The Samuel Roberts Noble Foundation
2510 Sam Noble Parkway, Ardmore, OK 73402, USA
Tel: 580 221 7391
Fax: 580 221 7380
Email: gdmay@noble.org
Insert length: 710
Plate: 016 Row: 2 Column: 08
Std Error: 0.00
Seq primer: TCACACAGGAAACGCTATGAC.

FEATURES
source
1..710
Location/Qualifiers
/organism="Medicago truncatula"
/db_xref="taxon:3880"
/clone="NF016F08FL"
/clone_1ib="Developing flower"
/tissue_type="Developing flowers"
/dev_stage="Developmentally pooled. Contains a mixture of very young, developing, fully-opened flowers and flowers in early transition into pods."
/note="Vector: Lambda Zap; cDNA was prepared from polyA+ enriched, pooled samples of equivalent amounts of total RNA from very young, developing, fully-opened flowers and flowers transitioning into pods. The cDNA was directionally ligated into the Uni-Zap XR vector (Stratagene) and packaged using the Gigapack III Gold packaging extracts. Phagemids containing cDNA inserts were in vivo excised from the recombinant Uni-Zap XR vector using Exassist helper phage and the E. coli strain XLI-Blue MRF (Stratagene). Excised plasmids were plated using SQR cells."

BASE COUNT
ORIGIN
219 a 205 c 94 g 187 t 5 others

Alignment Scores:
Pred. No.: 0.657 Length: 710
Score: 120.00 Matches: 26
Percent Similarity: 63.79% Conservative: 11
Best Local Similarity: 44.83% Mismatches: 13
Query Match: 7.19% Indels: 8
Gaps: 3

US-10-018-706-2 (1-322) x BI271940 (1-710)

QY 137 ProSeProVaAlaValaGlnSerSeraGProProVaAlaGlnHisProAlaValaGln 156
Db 293 CCTCTCCACCACTGTAATAATCA-----CCACCTTATCAATCAACCACTTATGTAAG 346
QY 157 LysProThProProVaAlaVala-----ValaValaLysProThProThr 171
Db 347 GCACCTTCTCCCTCTAGTGAACAACCACTCCACCTATTTGTGAATCACT--CCTTCT 403
QY 172 ProProVaValaGlnGlnProAlaProVaAlaProProVaAlaThGlnAlaPro 189
Db 404 CCTCTTGTGTGAACAACCACTTATCAATCAACCACTTATGTAAGGACACCA 457

RESULT 33
BM457264

LOCUS	BM457264	1460 bp	mRNA	linear	EST 05-FEB-2002
DEFINITION	AGENCOURT_6411711 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:5583548				
ACCESSION	BM457264				
VERSION	BM457264.1	GI:18506304			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
AUTHORS	1 (bases 1 to 1460)				
TITLE	NIH-MGC http://mgs.nci.nih.gov/.				
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)				
COMMENT	Unpublished (1992)				
	Contact: Robert Strausberg, Ph.D.				
	Email: cga@bbs-rt@mail.nih.gov				
	Tissue Procurement: ATCC				
	cDNA Library Preparation: Life Technologies, Inc.				
	DNA Sequencing by: Agencourt Bioscience Corporation				
	Clone distribution: MGC clone distribution information can be				
	found through the I.M.A.G.E. Consortium/LLNL at:				
	http://image.llnl.gov				
	Plate: L14M12346 row: m column: 21				
	High quality sequence start: 131				
	High quality sequence stop: 344.				
FEATURES	Location/Qualifiers				
source	1..1460				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
	/clone="IMAGE:5583548"				
	/clone_lib="NIH_MGC_92"				
	/tissue_type="embryonal carcinoma, cell line"				
	/lab_host="DH10B (phage-resistant)"				
	/note="Organ: testis; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 2.5 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library."				
BASE COUNT	233 a 669 c 208 g 350 t				
ORIGIN					
Alignment Scores:					
Pred. No.:	1.85	Length:	1460		
Score:	120.00	Matches:	36		
Percent Similarity:	45.38%	Conservative:	18		
Best Local Similarity:	30.25%	Mismatches:	33		
Query Match:	7.19%	Indels:	32		
DB:	13	Gaps:	7		
US-10-018-706-2 (1-322) x BM457264 (1-1460)					
QY	112 TrrpEuThrlEuTrrpSerGIYAspLeuysValArgGlarGserIleSerSerGIYal	131			
Db	432 TGGCTGCGGTGCACACGCGAACTGTCCAAAGCCGAAAAACCA-----	476			
QY	132 AenThralAhiSerHrProSerProValA-----ValGlnSerSerArg-----	146			
Db	477 -----CAADACCCCTGGCCCTCTCCCTCCGCGCCGTCCTCTCTCCACACCCCTTC	527			
QY	147 ProProValGlnGlnHisProAlaValGlnLysProHrProProValValValLys	166			
Db	528 CCCCTCCGACACAGTTCCACCGCGCCATGCGCTCCCGCCCGCCCA-----	575			
QY	167 LysProHrProHrProProProValValGlnGlnProAlaProValAlaProProValThr	186			
Db	576 -----ACACCCGCGCCCGCCCAATATTCACCCCGCCCGCTT-----CCCGCGTTCCT	626			
QY	187 GluAlaProPheAlaThrGlySerSerGIYalMet-----	198			
Db	627 AATCCCGCATTC-----TCCCGCAACTTCACACCTTCCCGCGCCCTTCCCGC	674			
QY	199 GlnPheArgIYrProValGlyAlaThrAsnProValValAlaGagPheGlyThrAla	217			

```

Db          675 CATTAACCGCTATCCCCGATACGGCCCTCCACTCATTACGCCGAGTTCACACATACC 731
RESULT 34
Bg607114
LOCUS
DEFINITION
WHE2489_E06_J11.S Triticum monococcum early reproductive apex cDNA
library Triticum monococcum cDNA WHE2489_E06_J11, mRNA
sequence.
ACCESSION
Bg607114
VERSION
Bg607114.1
KEYWORDS
SOURCE
ORGANISM
Triticum monococcum.
Triticum monococcum.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophytes; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
; Triticeae; Triticum.
1 (bases 1 to 567)
Anderson,O.D., Chao,S., Dubcovsky,J., Echenique,V., Han,P.S., Hsia
,C.C., Kang,Y., Lazo,G.R., Miller,R., Rauech,C.J., Seaton,C.L.,
Stamova,B. and Tong,J.C.
The structure and function of the expressed portion of the wheat
genomes - Early reproductive apex cDNA library from Triticum
monococcum
JOURNAL
Unpublished (2001)
COMMENT
Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacifica
West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105595773
Fax: 5105595818
Email: oanderson@pw.usda.gov
Sequence have been trimmed to remove vector sequence and low
quality sequence with phred score less than 20
Seq primer: Stragene SX primer.
FEATURES
Source
Location/Qualifiers
1..567
/organism="Triticum monococcum"
/cultivar="DV92"
/db_xref="taxon:4568"
/cclone="WHE2489_B06_J11"
/cclone_l1b="Triticum monococcum early reproductive apex
cDNA library"
/tissue_type="Early reproductive apex"
/dev_stage="Seven week-old plants"
/lab_host="E. coli XLORR"
/note="Vector: lambda Uni-ZAP XR, excised phagemid;
Site_1: EcoRI; Site_2: XhoI; The tissue, total RNA, and
poly(A) RNA were prepared from apex at double-ridge stage
to terminal-spikelet stage during transition from
vegetative state to flower state, a cDNA library was made,
and the cDNA clones were in vivo excised at the
University of California, Davis (V. Echenique, B. Stamova
, J. Dubcovsky). Plasmid DNA preparations and DNA
sequencing were performed in the OD Anderson lab (all
other authors)."
BASE COUNT      110 a      251 c      131 g      75 t
ORIGIN
Alignment Scores:
Pred. No.:           0.586              Length:           567
Score:               119.00             Matches:            46
Percent Similarity:  35.90%             Conservative:       10
Best local Similarity: 29.49%             Mismatches:         64
Query Match:        7.13%                 Indels:             36
DB:                  12                     Gaps:                5
US-10-018-706-2 (1-322) x BG607114 (1-567)
Oy          35 SerlysprrorhThrYraAsenSerThsencgLysergLyserHIsarphrsergLy-sergi 54
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db          36 TTCccgcgcggcgTHCAAGcctgcgccccaactgttcacctccgacgtaCaMaCcCaTccccca 95

```

QY 54 YGlyLeuAlaIleGlySerGlnValIleThrAspSerGlnGlyValProAsnArgTyrG 74
 DB 96 AGGTCTCAACCTCCGGCGTCAAGCCATC-----CCCCAGGTCTCCCGCCGGCGTCA 149
 QY 74 nValys---GlnGlyAspThrValSerIleAlaGlnArgTyrGlyLeuAsnTrpAr 93
 DB 150 ACCCGCCCGCCAGGTCTCAAGCCGGCGGTCAAGCCAGCCCGCCAGGTCTC----- 201
 QY 93 gGluIleGlyHisIleAsnLeuAsnSerTyrThrIleTyrThrGlyGlnTrpLe 113
 DB 201 ----- 201
 QY 113 uThrLeuTrpSerGlyAspLeuLysValArgGlnArgSerIleSerGlyValAsnTh 133
 DB 202 -----ACCGCGCGCGGTCAAGCCAGCCCGCCAGGTCTCAGCCGGCGTCAAGC 251
 QY 133 r-AlaHisThrProSerProValAla-----ValGlnSerSerArgProProValG 150
 DB 252 CTGTCCCGCAAGCCCTCACCGCGCGTCAAGCCCTGTCCCGCCAGCCCTCACCGCGCG 311
 QY 150 InGlnHisProAlaValGlnLysProThrProProValAlaValIleValLysProThrP 170
 DB 312 CCGCAACTCCCGCGCGCGCGTCAAGCCCGTCCACCG-----CCTTACAGGCCCGCGCG 365
 QY 170 rOThrProProValAlaGlnGlnProAlaProAlaProValAlaProPro 184
 DB 366 CTACTCTCTCAAGCTCAAGCCCGCCACCGCGCGCGCGCGCT 409

RESULT 35
 LOCUS B0646351 909 bp mRNA linear EST 15-JUN-2002
 DEFINITION AGENCOURT_8286354 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6299462
 5' mRNA sequence.
 ACCESSION B0646351 GI:21770523
 VERSION B0646351.1 GI:21770523
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 909)
 NIH_MGC http://mgi.mc.ncl.nih.gov/.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgaabs-r@mail.nih.gov
 Tissue Procurement: CGAP (Stanford)
 CDNA Library Preparation: Rubin Laboratory
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LMNL at:
 http://image.llnl.gov
 Plate: LRCM2512 row: c column: 15
 High quality sequence strip: 589.

FEATURES

source

Location/Qualifiers

1..909

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_11b="NIH_MGC_100"

/clone_11b="NIH_MGC_100"

/tissue_type="hepatocellular carcinoma, cell line"

/lab_host="DH10B (phage-resistant)"

/note="Organ: liver; Vector: pOTB7; Site: 1: XhoI; Site: 2:
EcoRI; cDNA made by oligo-dt priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
library."

BASE COUNT

179 a 332 c 235 g 163 t

ORIGIN

Alignment Scores:

Pred. No.:	1,16	Length:	909
Score:	119.00	Matches:	62
Percent Similarity:	37.77%	Conservative:	26
Best Local Similarity:	26.61%	Mismatches:	92
Query Match:	7.13%	Indels:	53
DB:	14	Gaps:	8

US-10-018-706-2 (1-322) x B0646351 (1-909)

QY 26 ThrThrCysIleLeuAlaGlyCysAlaSerLysProThrThrAsnSerThrSerGlySer 45
 DB 189 ACTTCTCAGAGACTTGAAGCTGACAGAGGTGTGTTACTTACCGCTCAACCTTCTC 248
 QY 46 -----GlySerHisArgThrSerGlySer--GlyGlyLeuAlaIleGly 59
 DB 249 TGCAGCCATCAACCGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 308
 QY 60 SerGlnValIleThrAspSer-GlnGlyValProAsnArgTyrGlnValLysGlnGlyAs 79
 DB 309 ACCCGCCAGCCAGCCAGCTTCAAGGAGCTGCGCGAGCCCTTGGCAGAGATCAAGGAGA 368
 QY 79 pThrValSerIleLeuAlaGlnArgTyrGly--LeuAsnTrpArgGluIleGlyHisIle 98
 DB 369 T-----GTTGGCAGAGCTTCAAGCTGATGATGATGATGATGATGATGATGATGATGAT 422
 QY 99 AsnAsnLeuAsnSerSerTyrThrIleTyrThrGlnGlnTrpLeu----- 113
 DB 423 ACGAAGCTCAAGAAAGATTGATGATGATGATGATGATGATGATGATGATGATGATGAT 482
 QY 114 ThrLeuTrpSerGlyAspLeuLysValArgGlnArgSerIleSerSerGlyValAsnThr 133
 DB 483 TCGCTCTGGGAGGCTTCAAGAGCTTCCGGAAGAGGCTCCAGAGGAGATGTTCTTC 542
 QY 134 AlaHisThrProSer-----ProValAlaValGln 143
 DB 543 TGTTCACAGAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 602
 QY 144 SerSerArgProProValGlnGlnHis----- 152
 DB 603 CCGCGGACACTCCCGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 662
 QY 153 -----ProAlaValGlnLysProThr-----ProProVa 162
 DB 663 CGCCCGCCCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 722
 QY 162 lValValValLysLysProThrProThrProProValAlaGlnGlnProAlaProValAl 182
 DB 723 CCGGATCTGACAGCGGCGCCAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 782
 QY 182 aProProValThrGlnAlaProPheAla--ThrGlySerSerGlyValMetGlnPheArgT 202
 DB 783 CCGACCTTCAACCCAGGCGCCCGCGCGCGCGCTTGGCGCGC-----CCCT 830
 QY 202 yProProValGlyAlaThrAsnProValVal 211
 DB 831 TCCCCCCCCCAGGAGCAACCCGTTCTC 859

RESULT 36

LOCUS A1844465

DEFINITION EST242475 tomato ovary, TAMU Lycopersicon esculentum cDNA clone

ACCESSION A1844465

VERSION A1844465.1 GI:4379836

KEYWORDS EST.

SOURCE tomato.

ORGANISM Lycopersicon esculentum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Asteridae; euasterids I; Solanales; Solanaceae; Solanum;

REFERENCE 1 (bases 1 to 443)
Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vialon,T., Holt,I.E.,
AUTHORS Liang,F., Upton,J., Roming,C.M., Craven,M.B., Fujii,C.Y., Bowman
,C.L., Nieman,W., Fraser,C.M., Venter,J.C., Martin,G.B., Tankeley
,S.D., and Giovannoni,J.
TITLE Generation of ESTs from tomato carpel tissue
JOURNAL Unpublished (1999)
COMMENT Other ESTs: EST242465
Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: <http://www.genome.clemson.edu/orders/index.html>.
Location/Qualifiers
1. .443
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="CLED1H16"
/clone_lib="tomato ovary, TAMU"
/tissue_type="carpel"
/dev_stage="5 days pre-anthesis to 5 days post-anthesis"
/lab_host="XLI-BIue MRP"
/note="Vector: pBluescript SK(-); Site 1: EcoRI; site 2:
XhoI; CLED - Tomato Carpel EST library. OligodT-primed and
directionally cloned cDNA in vector lambda ZAP II with 5'
and 3' ends located at the EcoRI and XhoI sites,
respectively."
BASE COUNT 119 a 187 c 31 g 106 t
ORIGIN
Alignment Scores:
Pred. No.: 0.507 Length: 443
Score: 118.00 Matches: 33
Percent Similarity: 48.19% Conservative: 7
Best Local Similarity: 39.76% Mismatches: 31
Query Match: 7.07% Indels: 12
Gaps: 5
US-10-018-706-2 (1-322) x A1484465 (1-443)
Qy 136 ThrProSerProValAlaValGlnSerSerArgProProValGlnGlnHisProAlaVal 155
Db 175 ACTCAACACACCACTATTGTC-----CATCCACAGTCACCCAAACCA----- 219
Qy 156 GlnHisProThrProProValVal-----ValValLysLysPro--ThrProThr 171
Db 220 CCATCACCACCACTCTATTGTTTACCCCCCATGTTTATCCACCAATCATCTCCACCA 279
Qy 172 ProProValValGlnGlnProAlaProProValAlaProProValThrGlnAlaProPheAla 191
Db 280 CCACTGTTGTGTACCTCCATCATCTTCAACACACCACTATTGTTCTCCACCTTTTGTCTC 339
Qy 192 ThrGlySerSerGlyValMetGlnPheArgTyr-----ProValGlyAlaThrAsn 208
Db 340 CCCAATCGTCGCGGTATATACACCACTTACGTCGCAAGTCCTGCGTTGTTACTCCA 399
Qy 209 ProValVal 211
Db 400 CCCATAGTT 408
RESULT 37
LOCUS BE460175 443 bp mRNA linear EST 18-MAY-2001
DEFINITION BE415467 tomato developing/immature green fruit Lycopersicon
esculentum cDNA clone cLEM9019, mRNA sequence.
ACCESSION BE460175
VERSION BE460175.1 GI:3504477
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
REFERENCE 1 (bases 1 to 443)
Alcala,J., Vrebalov,J., White,R., van der Hoeven,R.S., Holt,I.E.,
AUTHORS Liang,F., Hansen,T.S., Craven,M.B., Bowman,C.L., Roming,C.M.,
Nieman,W., Fraser,C.M., Martin,G.B., Giovannoni,J.J. and Tankeley
,S.D.
TITLE Generation of ESTs from tomato fruit tissue, immature green
JOURNAL Unpublished (2000)
COMMENT Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: <http://www.genome.clemson.edu/orders/index.html>
5 prime sequence.
Location/Qualifiers
1. .443
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cLEM9019"
/clone_lib="tomato developing/immature green fruit"
/tissue_type="fruit"
/dev_stage="immature green (5-35 days post-anthesis)"
/lab_host="SOLR"
/note="Vector: pBluescriptSKmuDapT; Site 1: EcoRI;
Site 2: XhoI; Fruit were tagged at 5 dpa (0.5 cm) and
harvested at 7 day intervals through 35 dpa. Equal masses
of tissue from each stage were combined (including seeds
and locules) prior to mRNA isolation."
BASE COUNT 117 a 194 c 30 g 102 t
ORIGIN
Alignment Scores:
Pred. No.: 0.507 Length: 443
Score: 118.00 Matches: 36
Percent Similarity: 44.35% Conservative: 19
Best Local Similarity: 29.03% Mismatches: 54
Query Match: 7.07% Indels: 15
Gaps: 5
US-10-018-706-2 (1-322) x BE460175 (1-443)
Qy 103 SerSerTyrThrIleTyrThrGlyGlnTrpLeuThrLeuTrp-----SerGly 118
Db 33 ACCCTCTATAGTGATACCTCCATGTTGTTCTCCACCTTCCACCACTTAACCAACCA 92
Qy 119 AspLeuLysValArgGluArgSerLysSerSerGlyValAsnThrAlaHisThrProSer 138
Db 93 AACACCACTTACCTCCAAACACACATACCAATCCACCTATGTTTACCCCTAT 152
Qy 139 ProValAlaValGlnSerSerArg-ProProValGlnGlnHisProAlaValGlnLys-- 157
Db 153 TGTATTACCAACATCATCTCAACACCACTATTGTCATCCACCACTCACTCAAAAC 212
Qy 158 -----ProThrProProValVal-----ValValLysLysPro--ThrProTh 171
Db 213 ACCATCACCAACACCTCTATTGTTTACCCCCCATGTTTATCCACCAATCATCTCCAC 272
Qy 171 rProProValValGlnGlnProAlaProProValAlaProProValThrGlnAlaProPheAla 191
Db 273 ACCACCTGTTGTGTACCTCCATCATCTTCAACACCACTATTGTTCTCCACCTTTTGT 332
Qy 191 ethGlySerSerGlyValMetGlnPheArgTyr-----ProValGlyAlaThrAs 208
Db 333 CCCCAATCTCTCCGTGTATATACCACTTACGTCGCAAGTCCTCGGTTGTTACTCC 392
Qy 208 nProValVal 211
Db 393 ACCCATAGTT 402
RESULT 38

H76194 615 bp mRNA linear EST 30-DEC-1997
 LOCUS 17899 Lambda-PRU2 Arabidopsis thaliana cDNA clone 196014T7, mRNA
 DEFINITION
 accession H76194 GI:1053445
 version H76194
 keywords
 source
 organism Arabidopsis thaliana
 reference Newman, T., deBruijn, F.J., Green, P., Keegstra, K., Kende, H., McIntosh, L., Ohlrogge, J., Raikhel, N., Somerville, S., Thomas, M., Retzel, E., and Somerville, C.
 title Genes galore: a summary of methods for accessing results from large-scale partial sequencing of anonymous Arabidopsis cDNA clones
 journal Plant Physiol. 106, 1241-1255 (1994)
 comment Contact: Thomas Newman
 MSU-DOE Plant Research Laboratory
 Michigan State University
 MSU-DOE-PRU, Michigan State University, Plant Biology Bldg., E. Lansing, MI
 Tel: 517-353-0854
 Fax: 517-353-9168
 Email: 223jlc@lhm.cl.msu.edu
 Seq primer: T7 dye primer.
 location/Qualifiers
 1..615
 /organism="Arabidopsis thaliana"
 /strain="var columbiana"
 /db_xref="taxon:3702"
 /clone="196014T7"
 /note="Vector: lambda Z1p-lox; Site 1: Sal; Site 2: Not; Lambda PRU2 is a cDNA library derived from equal quantities of 4 pools of mRNA. The mRNA sources were 1) 7 day germinated etiolated seedlings; 2) tissue culture grown roots; 3) staged plants half with 24 hour light cycle, half on 16 hr light; 8 hour dark-roseettes; 4) same plants as 3 but aerial tissue (stems, flowers and siliques). The vector is BRL's lambda Z1p-lox. The cDNA inserts were directionally cloned with Sal-Not arms using oligo dt primed cDNA. 182 t 24 others
 BASE COUNT 132 a 154 c 123 g 182 t 24 others
 ORIGIN
 Alignment Scores:
 Pred. No.: 0.902 Length: 615
 Score: 117.50 Matches: 36
 Percent Similarity: 42.37% Conservative: 14
 Best Local Similarity: 30.51% Mismatches: 50
 Query Match: 7.04% Indels: 18
 DB: 14 Gaps: 4
 US-10-018-706-2 (1-322) x H76194 (1-615)
 QY 134 AAlAhstHProSePrProValAlaValGInSeSerArGProProValGInGInHiPro 153
 Db 93 GCTCAAGCTCTGTCACACCAACACCGTACTCTCCACCAACGCGCTCTCTCA 152
 QY 154 ----AlaValGInLyPrProPro-----ProValValValValLyPrProTh 169
 Db 153 GTCACGACGAAACACCTTACCAATGCTTCACCGCTGTTCCAGTTAACAGCACT 212
 QY 170 ProPrProProValValGInGInPrProValAla----ProProValThrGInAla 188
 Db 213 CGAGCTCTACCACTTCCACACATCCCGGTTGCTTCTCTCCAGACCATGCT 272
 QY 189 ProPheAlaThrGInLySerGInLyVal-MetGInPheArGTYPrProValGInAlaThrAs 208

Db 273 CTTGCTCCCGGTCCTTCCTCGGTGATTAATCACTTCGTCACGCTTCAGGCGCCAG 332
 QY 208 nProValAlaArgArGheGlyThrAla----- 217
 Db 333 AACGCTCTCTATGATCAACAGTCCGCAATGGCTAACAAAGTTTCTCTTGTGGG 392
 QY 218 ----ThyValAlaGlySerThrValThrSerAsnGlyMetTrpPheSerGly 233
 Db 393 ACAAGCTGTCCGAGAGGTTTATACGGGATGCGTTTGGCTTAGAGG 444
 RESULT 39
 BH244723/c 754 bp DNA linear GSS 13-NOV-2001
 LOCUS ATZEC82TF ATZE Arabidopsis thaliana genomic clone ATZEC82, DNA
 DEFINITION
 accession BH244723 GI:16921883
 version BH244723
 keywords
 source
 organism Arabidopsis thaliana
 reference Arabidopsis thaliana
 eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
 1 (bases 1 to 754)
 Kaul, S., Town, C.D., Bowman, C.L., Van Aken, S.E., Uteback, T.V.,
 Feldblyum, T.V., and Fraser, C.M.
 Survey sequencing of Arabidopsis thaliana BAC T6316
 Unpublished (2001)
 contact: Chris Town
 TIGR
 9712 Medical Center Drive, Rockville, MD 20850, USA.
 Tel: 301-838-3523
 Fax: 301-838-0208
 Email: cdtown@tigr.org
 From Wash. U contig 1144. Caution: the DNA in this BAC may be from
 some non-Arabidopsis source
 Seq primer: TF
 Class: sheared ends.
 location/Qualifiers
 1..754
 /organism="Arabidopsis thaliana"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /clone="ATZEC82"
 /note="Vector: pHO52; Site 1: BstXI; 2-3 kb sheared BAC
 DNA inserted into pHO52 using BstXI linkers"
 BASE COUNT 124 a 199 c 355 g 76 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 1.21 Length: 754
 Score: 117.50 Matches: 32
 Percent Similarity: 48.28% Conservative: 10
 Best Local Similarity: 36.78% Mismatches: 40
 Query Match: 7.04% Indels: 5
 DB: 17 Gaps: 3
 US-10-018-706-2 (1-322) x BH244723 (1-754)
 QY 134 AAlAhstHProSePrProValAlaValGInSeSerArGProProValGInGInHiPro 153
 Db 402 GCTGCGCCGCGCGCGCGCTACTGTCACCTGCTGCCCCCATATATGTCGCGGCC 343
 QY 154 AAlaValGInLyPrProPro-----ValValValValLyPrProTh 171
 Db 342 GCATACACGCTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTAGCTGCGCA 283
 QY 172 ProPrProProValValGInGInPrProValAla----ProProValThrGInAla 191
 Db 282 GCCCATATGATGCG 229
 QY 192 ThrGInLySerGInLyValMetGInPheArGTYPrProValGInAlaThrAsnProValVal 211

Db 228 TCCTGCTCTCTCCGCGCGTCATCCACATACCCGTT---GCCCGAGCCCGATGATG 172
 Qy 212 ArgArgPheGlyThrAlaThr 218
 Db 171 CCCCGTACCCACCGCCACC 151

RESULT 40

U74096 329 bp mRNA linear EST 14-APR-1997
 LOCUS ATU74096 NaCl-treated Arabidopsis subtraction library Arabidopsis
 DEFINITION thaliana cDNA clone OS001, mRNA sequence.

U74096

U74096.1 GI:1932828

EST.

thale cress.

ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.

1 (bases 1 to 329)

Pih,K.T., Park,J.M., Jang,H.U., Kang,S.G., Piao,H.L. and Hwang,I.

EST of salt inducible mRNA in Arabidopsis thaliana

Unpublished (1997)

Contact: Kyeong Tae Pih

Department of Plant Molecular Biology

Gyeongsang National Univ., Plant Molecular Biology and

Biotechnology Research Center

Jinju, Kyounghnam 660-701, Korea

Tel: 82-591-751-5193

Fax: 82-591-759-9363

Email: inhwang@nongae.gsnu.ac.kr.

Location/Qualifiers

1..329

/organism="Arabidopsis thaliana"

/strain="Columbia"

/db_xref="taxon:3702"

/clone="OS001"

/clone_lib="NaCl-treated Arabidopsis subtraction library"

/dev_stage="7 day-old"

/note="Subtracted cDNA library from salt(NaCl)-treated

whole plants"

BASE COUNT 91 a 130 c 46 g 53 t 9 others

ORIGIN

Alignment Scores:

Pred. No.: 0.407

Score: 117.00

Percent Similarity: 51.72%

Best Local Similarity: 44.83%

Query Match: 7.01%

DB: 14

Gaps: 2

Length: 329

Matches: 26

Conservative: 4

Mismatches: 24

Indels: 4

US-10-018-706-2 (1-322) x U74096 (1-329)

Qy 131 ValAsnThrAlaHisThrProSerProValAlaValGlnSerSerArgProProValGln 150
 Db 53 GTGCTAACCGCGCAACGCCAATCCACCGCTCATTT-----ACACCAACCAACACNA 103
 Qy 151 GlnHisProAlaValGlnLysProThrProProValValValValLysLysProThrPro 170
 Db 104 ACTGCACCCGTCATTAACCAACCAACCAACCAACCAACCCGTCGTGACACCTCCACCAACCA 163
 Qy 171 ThrProProValValGlnGlnProAlaProValAlaPro---ProValThrGln 187
 Db 164 ACACCAACCCGTCATTAACCAACCAACCAACCAACCAACCTCTTACGCTTATTCAGAG 217

Search completed: July 3, 2003, 21:28:26
 Job time : 1155 secs

THIS PAGE BLANK (USPTO)